

PALEOCLIMATE DATA ASSIMILATION OVER THE COMMON ERA

by

Jonathan Michael King



Copyright © Jonathan Michael King 2022

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF GEOSCIENCES

In Partial Fulfillment of the Requirements  
For the Degree of

DOCTOR OF PHILOSOPHY  
WITH A MAJOR IN GEOSCIENCES

In the Graduate College

THE UNIVERSITY OF ARIZONA

2022

THE UNIVERSITY OF ARIZONA  
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Jonathan Michael King

titled: Paleoclimate Data Assimilation over the Common Era

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

*Kevin Anchukaitis*

Kevin Anchukaitis

Date: Apr 15, 2022

*Jessica Tierney*

Jessica Tierney

Date: Apr 15, 2022

*Connie Woodhouse*

Connie Woodhouse

Date: Apr 15, 2022

*Marcus Lofverstrom*

Marcus Lofverstrom

Date: Apr 15, 2022

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.



*Kevin Anchukaitis*

Kevin Anchukaitis

Department of Geosciences and School of Geography, Development, and Environment

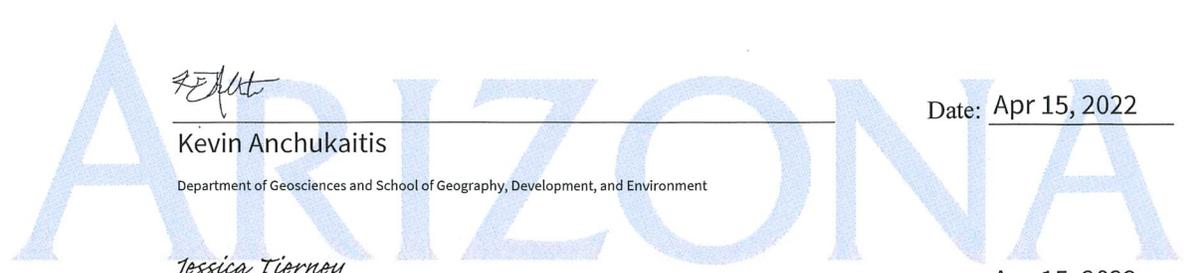
Date: Apr 15, 2022

*Jessica Tierney*

Jessica Tierney

Department of Geosciences

Date: Apr 15, 2022



## ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of many many people over the years. First off, an enormous thank you to Kevin Anchukaitis for his countless hours of advising, support, and mentoring. Kevin, I cannot thank you enough – this dissertation could not have happened without you, and I am truly grateful for your effort. Another huge thanks goes to Jessica Tierney for helping to shape my interests in data assimilation and for additional mentoring along the way. I also thank my committee members, Connie Woodhouse and Marcus Lofverstrom, for their thoughtful guidance over the years. Another major thank you goes to the many grad students and researchers in the climate systems center for their support and feedback over the years, especially Talia Anderson, Julie Edwards, Jessie Pearl, Kira Harris, Dervla Meegan-Kumar, Pablo Martinez-Sosa, Grace Windler, and Matt Osman. I'm sure none of them wanted to listen to a hundred talks about data assimilation, but they all did anyways, and this work benefited immensely from their feedback.

A special thanks to Matt Osman, Emily Judd, Jessica Badgeley, Brewster Malevich and everyone else involved in the testing and prototyping of the DASH toolbox. Coding is hard enough when everything works, but it takes special people to test out code that's definitely broken. Thanks for bearing with me, especially during those early stages.

Finally, all the thanks in the world to my family, partner, and friends. To my parents, for their tireless support over 23 years of school – I promise I'll get a real job now. To my brothers, for their perpetual readiness to plan an adventure, no matter how outlandish. To Lindsey, for all her love and support, for the best adventures, and for occasionally letting me win a board game night. Lastly, to all my friends in Tucson for sticking with me over the years – I couldn't have done it without you.

This research was made possible with grants from the National Science Foundation (AGS-1803946 and AGS-1602301), the Heising-Simons Foundation (grant 2016-05), the David and Lucinda Packard Foundation. Additional support was provided by a Galileo Circle Scholarship and the Data Science and Digital Scholarship Fellowship.

## DEDICATION

*To my parents, for their limitless love and support.*

## TABLE OF CONTENTS

LIST OF FIGURES . . . . .	8
LIST OF TABLES . . . . .	14
ABSTRACT . . . . .	16
CHAPTER 1 Introduction . . . . .	18
1.1 Background and Motivation . . . . .	18
1.2 Present Study . . . . .	23
1.2.1 Assimilating Northern Hemisphere temperatures with a small, highly-sensitive proxy network . . . . .	23
1.2.2 Assimilating the Southern Annular Mode . . . . .	25
1.2.3 DASH: A MATLAB toolbox for paleoclimate data assimilation . . . . .	27
1.3 References . . . . .	29
APPENDIX A A data assimilation approach to Last Millennium temperature field reconstruction using a limited high-sensitivity proxy network . . . . .	40
A.1 Abstract . . . . .	41
A.2 Introduction . . . . .	42
A.3 Methods . . . . .	46
A.3.1 Proxy Network . . . . .	46
A.3.2 Data Assimilation . . . . .	47
A.3.3 Pseudo-proxy Reconstructions . . . . .	49
A.3.4 Real NTREND Reconstruction . . . . .	52
A.4 Results . . . . .	54
A.4.1 Pseudo-proxy Experiments . . . . .	54
A.4.2 Real NTREND Reconstruction . . . . .	55
A.4.3 Epochal Temperature Changes . . . . .	57
A.4.4 Volcanic Response . . . . .	58
A.5 Discussion . . . . .	59
A.6 Conclusions . . . . .	63
A.7 References . . . . .	66
A.8 Supplemental Information . . . . .	78
A.8.1 Data Assimilation Methods . . . . .	78

TABLE OF CONTENTS – *Continued*

APPENDIX B Trends and variability in the Southern Annular Mode over the Com-	
mon Era . . . . .	106
B.1 Abstract . . . . .	107
B.2 Introduction . . . . .	108
B.3 Results . . . . .	113
B.4 Discussion . . . . .	115
B.4.1 Caveats and Limitations . . . . .	118
B.5 Conclusions . . . . .	119
B.6 Materials and Methods . . . . .	120
B.6.1 Southern Annular Mode Index . . . . .	120
B.6.2 Data Products . . . . .	121
B.6.3 Reconstruction Method . . . . .	123
B.6.4 Optimal Sensor Analysis . . . . .	128
B.6.5 External Forcing Analysis . . . . .	129
B.7 References . . . . .	131
APPENDIX C DASH: A MATLAB toolbox for paleoclimate data assimilation . .	160
C.1 Abstract . . . . .	161
C.2 Introduction . . . . .	162
C.3 Overview of Paleoclimate DA . . . . .	166
C.4 Description of DASH . . . . .	171
C.4.1 General Characteristics . . . . .	171
C.4.2 DASH Components . . . . .	172
C.5 Examples . . . . .	180
C.5.1 Northern Hemisphere Summer Temperatures over the Last Mil-	
lennium . . . . .	180
C.5.2 Global Sea Level Pressures at the Last Glacial Maximum . . . . .	186
C.5.3 Additional Considerations . . . . .	190
C.6 Warnings and Best Practices . . . . .	190
C.6.1 Temporal Variability . . . . .	191
C.6.2 Climate Model Biases . . . . .	192
C.6.3 Physically Inconsistent Reconstructions . . . . .	195
C.7 Discussion . . . . .	197
C.7.1 Future Development . . . . .	198
C.8 Conclusion . . . . .	199
C.9 References . . . . .	201
C.10 Supplementary Information . . . . .	213
C.10.1 Example 1: Northern Hemisphere Summer Temperatures over the	
Last Millennium . . . . .	213

TABLE OF CONTENTS – *Continued*

C.10.2 Example 2: Global Sea Level Pressures from the Last Glacial Maximum to Present . . . . .	220
--	-----

## LIST OF FIGURES

A.1	Locations of the 54 NTREND sites (Wilson et al., 2016). NTREND records were developed using ring-width data (TRW; circles), maximum latewood density (MXD; squares), or a mix of TRW, MXD, and blue intensity (Mixed; triangles). Marker color denotes the century in which each record begins. . . . .	89
A.2	Local Pearson’s correlation coefficients of pseudo-proxy reconstruction temperature anomalies with the target fields. Correlation coefficients are calculated over the period 850-1988 CE. Major rows indicate the model used to generate the target field, and major columns show the model used to build the initial ensemble for each assimilation. Minor rows designate whether the proxy network exhibits no time attrition or realistic time attrition. Minor columns indicate whether reconstructions use perfect or noisy proxies. The top-left and bottom-right quadrants display the perfect-model experiments, while the top-right and bottom-left quadrants show the biased-model cases. The black line in each map indicates 30°N.	90
A.3	Pseudo-proxy reconstruction skill for DA (left column), PPR (middle), and a comparison of the two (right). Skill metrics are relative to a CESM target field using noisy proxies and realistic temporal attrition. DA results are for a biased-model MPI prior. All skill metrics are computed over the period 850-1988 CE. In order the rows detail local Pearson’s correlation coefficients, RMSE values, temporal standard deviation ( $\sigma$ ) ratios, and mean biases. Comparison plots show DA skill minus PPR skill. The comparison plot of $\sigma$ ratios only considers grid points where $\sigma$ is underestimated in both the DA and PPR reconstruction. . . . .	91
A.4	Extratropical MJJA time series for the multi-model mean reconstruction (blue), Berkeley Earth instrumental records (yellow), and Anchukaitis et al. (2017) (red). We provide two different measures of uncertainty for the DA time series: the average of the $2\sigma$ posterior ensemble width taken over the 10 reconstruction (light grey), and the $2\sigma$ width of variability arising from prior model selection (dark grey). Reconstructed temperature anomalies are shown in Celsius for the instrumental era (top), and full reconstruction (middle). A three year moving average has been applied to the time series in the middle panel. The bottom panel displays the 31-year, running standard deviation of the DA ensemble-mean and Anchukaitis et al. (2017) time series. . . . .	92

LIST OF FIGURES – *Continued*

A.5	Spatial skill metrics for the multi-model mean reconstruction. Maps detail Pearson correlation coefficients (top left), RMSE values (top right), $\sigma$ ratios (bottom left), and mean biases (bottom right) of reconstructed grid point time series relative to the Berkeley Earth instrumental dataset over the period 1901-1988 CE. White markers show the proxy network and marker symbols follow the convention in Figure A.1. . . . .	93
A.6	Reconstructed temperature anomalies (in Celsius) between the MCA (950-1250 CE) and LIA (1450-1850 CE) for the DA reconstructions. Each map shows the results for a particular model prior. . . . .	94
A.7	As in A.6, but for the temperature CFRs summarized in Table A.3. . . .	95
A.8	Composite mean maps of the reconstructed temperature response in years containing a major tropical volcanic event. Events (N=20) are selected as tropical eruptions with a global forcing magnitude equal or larger than the 1884 Krakatoa eruption: this set consists of 916, 1108, 1171, 1191, 1230, 1258, 1276, 1286, 1345, 1453, 1458, 1595, 1601, 1641, 1695, 1809, 1815, 1832, 1836, and 1884 CE (Sigl et al., 2015; Anchukaitis et al., 2017). Temperature anomalies (in Celsius) are determined relative to the mean temperature of the five years preceding each volcanic event. Each map shows the results for a particular model prior. . . . .	96
A.9	As in Figure A.8, but for the temperature CFRs summarized in Table A.3 (rows). We only show grid points with reconstructed values for at least 6 eruptions. Maps show the composite mean response in years with a major tropical eruption (left), and in the year following a major eruption (right). . . . .	97
A.10	Spatial characteristics in the year following volcanic eruptions in 1257 (top) and 1600 (bottom) (De Silva and Zielinski, 1998; Lavigne et al., 2013) in the multi-model mean reconstruction. The left column displays temperature anomalies relative to the five preceding years in Celsius. The middle column shows the average $2\sigma$ width of the 10 posterior ensembles, and the right column shows the $2\sigma$ width of the multi-model ensemble. White markers show the proxy network for each event. Marker symbols follow the convention in Figure A.1. . . . .	98
A.11	As in Figure A.2, but for RMSE ( $^{\circ}\text{C}$ ). . . . .	99
A.12	As in Figure A.2, but for $\sigma$ ratios. . . . .	100
A.13	As in Figure A.2, but for mean biases ( $^{\circ}\text{C}$ ). . . . .	101

LIST OF FIGURES – *Continued*

A.14	Extratropical MJJA time series for the pseudo-proxy experiments with a CESM target. Reconstructed temperature anomalies are shown in Celsius (top) for the DA reconstruction (blue) and PPR reconstruction (red) along with the reconstruction target (yellow). The bottom panel displays a 31 year running standard deviation for each time series. A three year moving average has been applied to all time series. . . . .	102
A.15	As in Figure A.14, but for an MPI target. . . . .	103
A.16	As in Figure A.3, but for a MPI target field. Here, the DA reconstructions use a CESM prior. . . . .	104
A.17	Extratropical MJJA time series for the individual DA reconstructions. Each time series shows the results for a particular model prior. A 31 year moving average has been applied to each time series. . . . .	105
B.1	Map of the proxy network. Black Xs indicate the centroid of binned drought atlas sites. Grey markers indicate PAGES2k sites. The size of the PAGES2k markers correspond to the length of each record. Filled color contours show the field correlation between the 20CR SAM index with 20CR DJF sea level pressure over the period 1958-2000 CE. . . . .	146
B.2	Evolution of the reconstruction over time. Top: Comparison of the annual reconstruction (blue) with the Marshall index (red) over the instrumental era. Shading indicates the 5-95 percentiles of the reconstruction. Middle: Evolution of the annual reconstruction (blue) and 31-year lowpass filtered (black) over the Common Era. Shading indicates the 5-95 percentiles of the lowpass filtered series. Bottom: Composition of the proxy network over time. Colors for proxy types are as follows: Dark blue (coral), pink (ANZDA), grey (SADA), dark red (trees), light blue (glacier ice), black (lake sediment). . . . .	147
B.3	Optimal Sensor Analysis. Top: Maps of the potential ability for drought atlas (left) and PAGES2k (right) sites to constrain reconstruction posterior variance. Middle: Evolution of reconstruction posterior variance over time. Yellow bars indicate the addition of the indicated proxy to the network. Bottom: Ranked histograms of the seven sites with greatest potential influence in the early reconstruction (left), immediately before the addition of drought atlases (middle) and for the full network (right). Potential influence is determined as the uncertainty constrained by a single-proxy network. . . . .	148

LIST OF FIGURES – *Continued*

B.4	SAM climate responses. Top left: Wavelet coherence of the reconstructed SAM index with the solar forcing reconstruction (Schmidt et al., 2012). Bottom left: Composite mean response to major volcanic eruptions. Shading indicates 5-95 percentiles. Blue line is the ensemble mean. Right: Instrumental SAM trends for the Marshall Index (top right) and reconstruction (bottom right). Colored points indicate trends calculated from a sliding window centered on the given year. Solid (dotted) contours surround statistically significant trends at the 90% confidence interval relative to the reconstruction over the period 1500-1900 CE (1-1900 CE). . . . .	149
B.5	Comparison of SAM reconstructions over the last millennium. All reconstructions are smoothed via a 30-year Gaussian filter and normalized to the period 1400-1850 CE. . . . .	150
B.6	Trend analysis as in Figure B.4, but using the early reconstruction (1-899 CE) to quantify natural variability and assess trend significance. . . . .	151
B.7	Correlation of reconstructions with the Marshall Index for different latitude cutoff boundaries. Correlation is assessed over the period 1958-2000 CE. Blue line shows results for reconstructions with an MME prior constructed from all 10 models. Red line shows results for reconstructions with an MME prior built from the 4 high-resolution models. . . . .	152
B.8	Skill comparison for different priors. Skill metrics are assessed relative to the Marshall Index over the period 1958-2000. Prior to skill assessment, reconstruction time series are scaled such that the mean and standard deviation of the detrended reconstructions match the mean and standard deviation of the detrended Marshall Index. Black circles indicate values for the all-model MME prior (10 models). Red triangles indicate values for the high-resolution MME prior (4 models). Figure columns are 1. Correlation, 2. Root mean-square error, 3. Standard deviation ratios (computed as the ratio of reconstruction variability over Marshall Index variability, and 4. Bias in the mean value of the series. To condense the plot, negative mean biases are shown here. . . . .	153

LIST OF FIGURES – *Continued*

- B.9 Variance correction scheme. Top: Raw assimilation (red) and the variance-adjusted final reconstruction (blue). The raw assimilation displays less variance than the final reconstruction, particularly during the early part of the record. Middle: Solid lines show the moving 30-year standard deviations of the raw assimilation (red) and variance-adjusted reconstruction (blue). Dotted lines indicate the linear trends in the moving standard deviation time series over the full period. Bottom: The multiplicative scaling weights used to produce the final, variance-adjusted reconstruction. Weights are applied multiplicatively to the raw assimilation in each time step. . . . . 154
- C.1 Illustration of common tasks and vocabulary for paleoclimate data assimilation. Top left: Gridded climate model output is reshaped into a *state vector*. Red triangles indicate the locations of proxy records. Top right: Multiple climate model outputs are reshaped into state vectors and concatenated into an *ensemble*. Bottom left: Forward models are applied to each state vector and used to generate *proxy estimates* for each proxy record. Bottom right: Proxy estimates are compared directly to the real proxy records. The difference between the estimates and the real records is the *innovation*. . . . . 226
- C.2 Flowchart illustrating DASH components and their uses within the context of paleoclimate data assimilation workflows. . . . . 227
- C.3 Results from Example 1, the NTREND assimilation. Top: Reconstructed mean extratropical summer (June-August) temperatures. The blue line shows the reconstructed index when the index is assimilated directly in the state vector. The red line shows the index calculated from the posterior spatial field. Grey shading indicates the 5-95% confidence level for Index 1. Lower left: The reconstructed summer-temperature spatial field in the year 1850 CE. Lower center: The variance of the posterior spatial field in the year 1850 CE. High variance indicates greater uncertainty in the reconstructed spatial field. Lower right: Results of the optimal sensor analysis. Circles indicate the locations of the NTREND tree-ring records. The color of each circle indicates the percent variance of the reconstructed index that is constrained by assimilating each NTREND site individually. 228

LIST OF FIGURES – *Continued*

- C.4 Results from Example 2, the LGM assimilation. Upper row shows results for the Last Glacial Maximum (18-21 ka); lower row shows results for the most recent 3,000-year interval (0-3 ka). From left to right, columns display reconstructed sea level pressure fields (hPa), the standard deviation across the posterior ensembles for each reconstructed field (hPa), and the percent errors from the proxy validations. . . . . 229

## LIST OF TABLES

A.1	Summary of climate models used to construct data assimilation prior ensembles. Climate models are listed along with the identifying acronym used in this study. The years of available output are provided with the experiment used to generate them. The size of the model prior generated from these years is also provided. Taylor et al. (2012) provide more details on the PMIP3 and CMIP5 experiments, and Otto-Bliesner et al. (2016) describe the LME. . . . .	82
A.2	Calibrated localization radii. Localization radii for individual model priors are selected using the radius search and calibration-validation procedure detailed in section A.8.1. Skill metrics are the median values obtained for the mean extratropical MJJA time series relative to BEST for the set of validation periods. . . . .	83
A.3	Temperature field reconstructions used to compare spatial patterns of climate response to radiative forcings in this study. We provide a reference for each CFR along with the name used in this study. We also note the maximum size of the proxy network used in each study along with the target temperature fields. . . . .	84
A.4	Withheld proxy verification statistics for individual models. Reported skill metrics are the median for all individual proxy comparisons over the 54 leave-one-out assimilations. . . . .	85
A.5	As in Table A.2, but using the RMSE optimization scheme. . . . .	86
A.6	Pseudo-proxy localization radii and split-sample validation metrics. As in Table A.2, but using climate model output as the target field. . . . .	87
A.7	Skill metrics for pseudo-proxy reconstructions of mean extratropical May-August time series. DA reconstructions use the realistic biased-model, noisy-proxy, time-attrition experimental design. PPR time series and target time series are calculated using only the grid cells for which $RE > 0$ in each reconstructed time step. . . . .	88
B.1	Reconstruction Skill . . . . .	145
B.2	Climate models tested for use in assimilation priors. . . . .	155
B.3	Latitudes with the most strongly anticorrelated zonal-mean SLP anomalies in tested climate models. Correlation coefficients are calculated for DJF seasonal means over all available years in the interval 850-2005 CE. . . . .	156
B.4	PAGES2k Sites used in the final reconstruction. The included sites are south of 25°S, and have annual or higher resolution. . . . .	159

LIST OF TABLES – *Continued*

C.1	Proxy forward-models currently supported by DASH. . . . .	225
-----	---	-----

## ABSTRACT

Past climates provide key insights into the drivers and behavior of the Earth's climate system, and such insight is highly valuable in the context of anthropogenic climate change. The Common Era is a particularly useful period for scientific inquiry because it represents the baseline climate prior to the industrial revolution and provides a comparison point for understanding observed climates. Climate reconstructions, in which climate proxy records are leveraged to estimate past climate fields and variability, are a powerful tool for studying past climates. Recently, data assimilation (DA) has emerged as a promising reconstruction technique. Unlike traditional reconstruction methods, DA integrates climate proxy records with climate model output, thereby leveraging strengths of both information sources. This dissertation presents three studies that develop DA reconstruction methodologies and leverage DA for Common Era climate reconstructions. The first study examines the consequences of applying DA to a small, highly sensitive proxy network. We find that the method underestimates temporal variance as the proxy network becomes sparse. We also observe that DA is sensitive to climate model biases but find that the use of multi-model ensembles helps retain reconstruction skill. Using these insights, we then use DA to reconstruct summer temperatures in the Northern Hemisphere spatially over the last millennium. The second study uses a DA framework to reconstruct the Southern Annular Mode, a major mode of climate variability. We optimize the assimilation using insights from the first study, and also extend DA methodologies to accommodate gridded, spatially-covarying drought atlases as proxy records. We also adapt DA into an optimal sensor framework, which allows us to quantify the influence of individual proxy records on the reconstruction. Our reconstruction reduces major uncertainties inherent in existing SAM reconstructions and extends those reconstructions by a full millennium. In the final study, we present DASH, a MATLAB toolbox designed to facilitate paleoclimate data assimilation. This toolbox is motivated by the practical difficulty of implementing DA

methods for generalized paleoclimate analyses and provides command-line and scripting routines that implement common tasks in DA workflows. DASH is highly modular so supports paleoclimate DA for a diverse range of time periods, spatial regions, proxy networks, and algorithms. This toolbox consolidates and clarifies complex DA methods and can serve as a tool for paleoclimatologists with varying areas of expertise. Overall, these papers help develop and establish DA as powerful tool for paleoclimate reconstruction.

## CHAPTER 1

### Introduction

#### 1.1 Background and Motivation

Past climates provide valuable insight into the drivers, variability, and evolution of the climate system, and these insights are particularly relevant in the context of ongoing anthropogenic climate change (Alley, 2003; Hargreaves et al., 2007; Rice et al., 2009; Schmidt, 2010; Snyder, 2010; Bell et al., 2013; Ault et al., 2014; Coats et al., 2020; Tierney et al., 2020a). Past climates can serve as analogues for future warm climate states (Overpeck et al., 2006; Burke et al., 2018; Tierney et al., 2020a), act as comparison points for modern climates (see Gulev et al. (2021) and references therein), and quantify natural and forced climate variability (Cane et al., 2006; Cook et al., 2011; Goosse et al., 2012a; Ault et al., 2013; Fernández-Donado et al., 2013; Neukom et al., 2019a; Fang et al., 2021). By providing targets for climate model skill assessments, past climates help improve climate forecasts and thereby enable superior adaptation strategies (Crowley, 1991; Hargreaves and Annan, 2009; Schmidt et al., 2013; Zhu et al., 2021a,b; Gulev et al., 2021). The Common Era (i.e. the last two millennia) is a particularly useful period for paleoclimatic inquiry. The continental configuration and orbital characteristics of the Earth system over the Common Era closely match modern values; thus, the Common Era can help quantify the natural climate variability of the contemporary Earth system and provides a comparison point for post-industrial climate states (e.g. PAGES 2k-PMIP3 group, 2015; Otto-Bliesner et al., 2016). The Common Era also coincides with the development of all major modern societies and represents a climate state that is known to support complex human civilizations. Finally, because much of it immediately precedes the industrial revolution, the Common Era provides a baseline comparison point for modern climates (see Gulev et al., 2021).

Although human civilizations have existed throughout the Common Era, there are few

instrumental climate records prior to the twentieth century, so scientists must rely on alternate information sources to study the climate during this period. Traditionally, paleoclimatologists have leveraged two tools to study past climates: (1) climate proxy records and associated climate reconstructions generated using statistical estimation techniques, and (2) climate model simulations run with temporally appropriate climate-system boundary conditions and forcings. Both methods have advantages and tradeoffs for studying past climates. A major advantage of proxy reconstructions is that they estimate the actual historical trajectory of the climate system. However, these reconstructions are hindered by a number of factors, including non-climatic noise in the proxy records and time-uncertainty when establishing proxy chronology. Additionally, many reconstructions target climate metrics or indices, and so may not illustrate the full-field state of the climate system. Furthermore, proxies record local climate conditions and often a specific season; thus, proxies do not always give a complete spatiotemporal picture of prevailing climate conditions. Many reconstructions target climate variables that are not directly sensed by the proxy records, and so must establish the relationships between proxies and reconstruction targets using empirical (often linear) statistical approaches or proxy system modeling. However, the physical processes leading to the archival of climate signals in proxy records are often complex or not fully understood, which complicates both interpretation and attempt to model these systems. Similarly, non-stationarity in the climate system potentially hinders most statistical approaches. For example, many reconstructions calibrate proxy records to the instrumental period, but the modern climate system is both sparsely observed and already responding to anthropogenic climate change. The long-term true expression of climate system covariances may therefore differ from those of the modern period. Finally, many proxy reconstructions that target circulation indices implicitly assume that teleconnections between local- and large-scale climate variables are stationary over reconstruction periods, an assumption that often does not hold in reality.

By contrast, paleoclimate model simulations provide information for all climate variables and spatial points in the model domain. Furthermore, modeled variables evolve according to physical governing equations and parameterizations, rather than the (often linear) statistical assumptions common in proxy reconstructions. Thus, climate model

simulations are advantageous for investigating the physical mechanisms underlying the behavior and evolution of past climate states. However, no model can capture the full complexity of the Earth system, and so all simulations necessarily contain errors in their representation of past climates. The spatial and temporal resolutions of models are finite, and so processes that occur on scales below model resolution must be estimated using parameterizations, rather than the direct evolution of simulated climate variables. Models are typically tuned to the observed climate of the late twentieth century, but these tunings are not necessarily representative of other time periods, particularly over geological time scales. Finally, many features of the climate system are either chaotic or dominated by internal variability over a range of time periods (Deser et al., 2012). Paleoclimate simulations are therefore extremely unlikely to represent the actual historical trajectory of the climate system, but rather a single trajectory from a distribution of possible past climate states (e.g. Kay et al., 2015).

Recently, paleoclimate data assimilation (DA) has emerged as a novel tool for studying past climates (e.g. Bhend et al., 2012; Goosse et al., 2012b; Mairesse et al., 2013; Hakim et al., 2016; Steiger et al., 2017, 2018; Tardif et al., 2019; Tierney et al., 2020b; King et al., 2021; Osman et al., 2021; Zhu et al., 2022; King et al., 2022). Assimilation methods integrate the climate signals recorded in proxy records with dynamical constraints provided by climate models. In doing so, they combine the strengths of both information sources. For example, leveraging climate models allows assimilations to reconstruct full spatial fields, and reconstructed variables are initialized via the model's physical governing equations. Simultaneously, the climate model states are constrained by proxy records and so reconstructed variables should reflect the actual historical trajectory of the climate system. DA methods relax the assumption that proxy records and distal climate variables exhibit a stationary relationship over the instrumental era. Instead, these methods typically only assume the stationarity of proxy response to local climate variables. Also, because assimilation techniques use proxy (system) forward models to describe the relationship between proxy records and local climate, DA supports nonlinear and physically realistic proxy responses. Finally, DA methods do not require stationary climate system covariances, because changes to the covariances can be captured by the

climate model's response to changing boundary conditions and forcings.

There are several paleoclimate DA paradigms, but most recent efforts have focused on ensemble assimilation methods, namely ensemble Kalman filters (EnKF) and particle filters. In this dissertation, I have focused on EnKF, which has been shown to perform well compared to other paleoclimate DA methods (Liu et al., 2017), and which has recently been used to generate Common Era reconstructions (e.g. Hakim et al., 2016; Tardif et al., 2019; Steiger et al., 2018; Franke et al., 2020). EnKF methods proceed by updating an ensemble of climate states to more closely match a set of paleoclimate proxy records. There is no strict definition of a climate state, but they typically include one or more climate variables at a set of spatial points over some period of time. These states are produced using one of two approaches. In the online approach, states are selected from a time-evolving set of parallel model simulations whose state variables are iteratively constrained by the assimilation of the proxy records (e.g. Perkins and Hakim, 2017). In the offline approach (Oke et al., 2002; Evensen, 2003), climate states are generated from pre-existing climate model output and the proxies do not inform the model's evolution (e.g. Bhend et al., 2012; Annan and Hargreaves, 2012; Steiger et al., 2014; Hakim et al., 2016; Valler et al., 2019; Tierney et al., 2020b; King et al., 2021; Osman et al., 2021; Tardif et al., 2019; Franke et al., 2020). In this dissertation, I have focused on the offline method because (1) offline regimes have been found to perform similarly to online methods in paleoclimate contexts at a fraction of the computational cost (Matsikaris et al., 2015; Acevedo et al., 2017), and (2) the offline paradigm is currently the more prevalent method for paleoclimate reconstructions (compare Perkins and Hakim (2017) with Bhend et al. (2012); Goosse et al. (2012b); Mairesse et al. (2013); Hakim et al. (2016); Steiger et al. (2017, 2018); Tardif et al. (2019); Tierney et al. (2020b); King et al. (2021); Osman et al. (2021); Zhu et al. (2022); King et al. (2022)).

An offline EnKF assimilation begins by collecting climate model output into an ensemble of climate states. We refer to this ensemble as the *prior*, and it provides an initial distribution of possible values for a past climate. The algorithm proceeds by comparing the climate states in the prior to the proxy records and updating the climate states to more closely match the proxies. The ensemble of updated climate states is known as

the *posterior* and represents the constrained model output. Typically, the mean of the climate states in the posterior is used as a reconstruction, and the spread of values across the posterior quantifies reconstruction uncertainty. In order to compare the climate states with the proxy records, EnKF must move the climate states and proxies into a common unit space. This is accomplished by running proxy (system) forward models on the climate states. The forward models simulate proxy records for each climate state, known as *proxy estimates*, and these estimates allow direct comparison of the climate states with the real proxy records. The difference between the proxy records and estimates is known as the *innovation* and is used to update the ensemble of climate states. The innovation is weighted by proxy covariances and uncertainties, and then propagated through an estimate of climate system covariance derived from the proxy estimates and prior. This process produces an update for each climate state in the prior, which is used to generate the posterior ensemble.

Despite the potential of paleoclimate DA, these methods have so far seen relatively modest use in the reconstruction literature. This is in part a consequence of the relative novelty of DA as a reconstruction method. Paleoclimate applications present unique challenges for assimilation frameworks, and much work has focused on developing DA algorithms for diverse paleoclimate contexts (e.g. Bhend et al., 2012; Dirren and Hakim, 2005a,b; Steiger et al., 2014; Mairesse et al., 2013; Dubinkina et al., 2011; Franke et al., 2020; Dee et al., 2016; King et al., 2021; Parsons et al., 2021; Tierney et al., 2020b; Osman et al., 2021). Additionally, DA methods remain difficult to implement in practice due to their multi-faceted requirements and limited software support. This dissertation contributes to the ongoing development of paleoclimate DA frameworks by presenting three studies that extend DA reconstruction methodologies, leverage DA for Common Era climate reconstructions, and provide software tools facilitating DA for generalized paleoclimate analyses.

## 1.2 Present Study

This dissertation focuses on the development of paleoclimate data assimilation methodology and the subsequent reconstruction of Common Era climate variables. This work is divided into three manuscripts that are included as Appendices to the dissertation. We first investigate the consequences of assimilating a small network of highly-sensitive tree ring records in the context of last millennium temperature and use the insights from the analysis to reconstruct summer temperatures over the extratropical Northern Hemisphere for the last millennium (Appendix A). Key insights from this study include the tendency for EnKF to underestimate a reconstruction's temporal variance, and the ability of multi-model climate model ensembles to improve reconstruction skill. Informed by this first study, we next use EnKF methodology to reconstruct the Southern Annular Mode (SAM) at annual resolution over the Common Era (Appendix B). The reconstruction extends SAM reconstructions by a millennium and the DA method reduces major uncertainties inherent to previously published reconstructions. In addition to the SAM reconstruction, Appendix B also includes several new developments to DA methodology, including (1) extending EnKF methods to assimilate gridded, spatially-covarying drought atlases, and (2) adapting EnKF into an optimal sensor framework in order to assess the influence of individual proxy records on the final reconstruction. Finally, in Appendix C, we present a MATLAB toolbox designed to facilitate paleoclimate data assimilation. This toolbox implemented the analyses presented in Appendices A and B but is not based on any particular reconstruction. Instead, the package is designed to facilitate general assimilation workflows.

### 1.2.1 Assimilating Northern Hemisphere temperatures with a small, highly-sensitive proxy network

In Appendix A, we examine the effects of using a small network of highly sensitive temperature proxies to assimilate Common Era surface temperature and use the insights gleaned from this investigation to reconstruct summer temperatures in the extratropical Northern Hemisphere over the last millennium. When selecting a proxy network, re-

constructions must balance the trade-offs of network size and climate sensitivity (Esper et al., 2005; Frank et al., 2010; Wang et al., 2015; Wilson et al., 2016; Anchukaitis et al., 2017; Esper et al., 2018; Franke et al., 2020; Cort et al., 2021). Large networks provide greater spatial coverage, but the records in these networks are often less strictly screened for climate sensitivity. Thus, large networks may contain records only weakly sensitive to reconstruction targets, which can reduce reconstruction skill. By contrast, the proxy records in small, curated networks typically exhibit higher climate sensitivity, at the cost of reduced coverage. Many Common Era temperature assimilations leverage large networks, typically derived from the global PAGES2k network of temperature-sensitive proxy records (e.g. Hakim et al., 2016; Tardif et al., 2019; Zhu et al., 2020). However, work by Franke et al. (2020) suggests that smaller proxy networks with greater climate sensitivity can outperform large networks in temperature assimilation contexts. Despite this evidence, there are few small-network assimilations in the paleoclimate literature, and so the effects of these networks on assimilated reconstructions are poorly characterized.

To address this uncertainty, we use the Northern Hemisphere Tree-Ring Network Development (NTREND) network (Wilson et al., 2016; Anchukaitis et al., 2017) to examine the effects of assimilating a small, highly-sensitive proxy network for last millennium temperature reconstructions. The NTREND network consists of 54 curated, temperature-sensitive tree ring records, and represents a smaller, more sensitive alternative to the PAGES2k network. The study begins with a series of pseudo-proxy experiments (Mann and Rutherford, 2002; Zorita et al., 2003; Smerdon, 2012) designed to test the sensitivity and skill of the DA reconstruction method. These pseudo-proxy experiments use simulated proxy records generated from last millennium climate model simulations in order to examine reconstruction properties in the context of a known target. We find the assimilation method is highly sensitive to climate model covariance biases, which can reduce skill far more strongly than noise in the proxy records or proxy network attrition. The assimilation method also severely underestimates the temporal variance of reconstructed temperatures, an effect that becomes more pronounced with reduced network size. Consequently, network attrition causes the early reconstruction to exhibit sharply lower temporal variance than the later period. Nonetheless, the pseudo-proxy reconstructions exhibit high

skill in the extratropical Northern Hemisphere and demonstrate the potential to improve on existing temperature reconstructions.

With this context, we then produce a real reconstruction using the observed NTREND network. To help minimize the effects of climate model covariance biases, we use a multi-model ensemble of 10 assimilations. Each ensemble member uses output from a different climate model, and the multi-model mean demonstrates greater skill than any single member alone. We also quantify spatial and temporal uncertainties throughout the reconstruction and identify tree-line North America and eastern Siberia as regions that would particularly benefit from additional proxy development. Finally, we compare our ensemble-mean temperature reconstruction to existing last millennium temperature reconstructions (Anchukaitis et al., 2017; Guillet et al., 2017; Zhu et al., 2020; Tardif et al., 2019; Neukom et al., 2019b). We find substantial differences across all products and emphasize the need to better quantify the effects of network selection and methodological choices in future reconstruction products.

In addition to producing a new field reconstruction of last millennium temperatures, this study emphasizes several key points regarding the implementation of DA methods for Common Era reconstructions. In particular, we note the reduction of temporal variance, utility of multi-model ensembles, and the need to quantify the effects of network composition as topics addressed in Appendices B and C.

### **1.2.2 Assimilating the Southern Annular Mode**

In Appendix B, we use data assimilation to reconstruct the Southern Annular Mode (SAM) over the Common Era. The SAM is the dominant mode of extratropical atmospheric variability in the Southern Hemisphere, and its variations affect societies and ecosystems throughout the globe (Rogers and Van Loon, 1982; Gong and Wang, 1999; Thompson and Wallace, 2000; Marshall, 2003; Thompson and Solomon, 2002; Kwok and Comiso, 2002; Gillett et al., 2006; Gupta and England, 2006; Marshall et al., 2006; Hendon et al., 2007; Van Lipzig et al., 2008). Despite its importance, the SAM's behavior is poorly constrained over the Common Era, and existing reconstructions exhibit major differences prior to the 1900s (compare Abram et al., 2014; Villalba et al., 2012; Dätwyler

et al., 2018). A possible cause of this discrepancy regards the calibration of existing reconstructions against an instrumental SAM index. These calibrations are used to relate proxy records to the SAM reconstruction target, but they assume that the SAM's teleconnections remain stationary with respect to the climate variables sensed by the proxy records. However, there is strong evidence that the SAM's teleconnections are not stationary over the instrumental period, so this assumption is a major source of uncertainty for existing reconstructions (Silvestri and Vera, 2009; Gallant et al., 2013; Hessel et al., 2017). Further differences may arise from reconstruction's use of different proxy networks.

The nature of these uncertainties recommends DA as an ideal method for reconstructing the SAM. Unlike existing efforts, DA does not rely on direct calibration with reconstruction targets; instead, calibration proceeds relative to the local climate variables sensed by the proxy records. Consequently, DA reconstructions of the SAM are not sensitive to non-stationary teleconnections. Additionally, DA methods can be adapted into uncertainty quantification frameworks, which provides insight into the effects of different proxy networks. As a consequence of these considerations, we use this study to develop DA methodologies for reconstructing the SAM.

This study includes several developments to DA methodology, many prompted by insights from Appendix A. To minimize the effects of climate model biases on the reconstruction, we utilize a multi-model prior. To account for the effects of offline DA on reconstructed temporal variance, we develop a novel variance correction method that accounts for changes to assimilated proxy networks. Finally, to better quantify the effects of proxy network composition on the final reconstruction, we adapt our DA method into an optimal sensor framework. This optimal sensor method allows us to quantify the relative influence of individual proxies on the reconstruction at any given point in time, and represents the first such use of optimal sensors in paleoclimate DA. In addition to developments prompted by Appendix A, we also develop methodology for assimilating gridded, spatially-covarying proxy records. This allows us to include two drought atlases in our assimilation, the South America Drought Atlas (SADA; Morales et al., 2020), and the Australia-New Zealand Drought Atlas (ANZDA; Palmer et al., 2015). The inclusion of these drought atlases in our network represents a significant increase in the amount of

paleoclimate information available to our reconstruction, potentially improving the final product.

In this study, we produce the first SAM reconstruction spanning the full Common Era at annual resolution. We generate the reconstruction by using EnKF to integrate the PAGES2k network (PAGES2k Consortium, 2017), SADA, and ANZDA with a multi-model prior constructed from four high-resolution coupled climate models. The reconstruction exhibits high skill relative to instrumental SAM indices, with skill values comparable to existing reconstructions. Because our DA method is not calibrated to an instrumental SAM index, it is not sensitive to the assumptions of SAM teleconnection stationarity that have hindered previous reconstructions. Using the optimal sensor framework, we find that the reconstruction is most strongly influenced by the drought atlases, and by tree-ring records from New Zealand and Tasmania. Several ice cores from central Antarctica also display high influence during the first millennium of the reconstruction.

Our reconstruction provides a foundation for investigating the SAM's behavior over the Common Era. In this study, we use wavelet and superposed-epoch analyses to investigate the response of the SAM to solar and volcanic forcings. Overall, we find the SAM minimally affected by these forcings and conclude that the SAM's variations represent internal variability over the pre-industrial Common Era. By contrast, trend analysis indicates that the modern multidecadal trends in the SAM are outside the range of variability experienced over the last two millennia. This finding places recent SAM trends in a long-term context and supports evidence that the SAM's recent variations represent a response to anthropogenic forcings.

### **1.2.3 DASH: A MATLAB toolbox for paleoclimate data assimilation**

In Appendix C, we present DASH, a MATLAB toolbox designed to facilitate paleoclimate assimilation workflows. The creation of DASH is motivated by the practical difficulty of implementing paleoclimate data assimilation. Assimilation studies must integrate data from multiple scientific realms, including climate model output, proxy records, forward models, and uncertainty assessments, which are often stored in different data formats, across multiple files, and with different characteristic metadata. Even once data are or-

ganized into a common format, many computational tasks are required before running a DA algorithm; these tasks include designing state vectors, running forward models, and quantifying record uncertainties. Throughout all these tasks, scientists must exercise care to ensure that reconstruction targets, proxy model inputs, and proxy estimates remain aligned within a state vector ensemble, which can be difficult to implement. After these tasks are complete, scientists must design a DA algorithm, considering factors such as state vector design, covariance localization, use of evolving priors, and choice of filter. Finally, from a technical standpoint, efficiently implementing an ensemble DA requires detailed knowledge of both the mathematics underlying the filter, and the computational efficiencies inherent to a given coding language.

As a result of these complexities, paleoclimate DA can be difficult to implement. For example, a scientist with a background in climate modeling may not have experience running proxy system models, while a proxy system expert may not have sufficient knowledge of ensemble DA algorithms to design a computationally efficient filter. Although several paleoclimate DA codes do exist, these are built around specific analyses and are difficult to adapt to new DA studies. DASH is designed to alleviate these difficulties by providing intuitive and simple commands to automate common DA tasks. The toolbox is written in an object-oriented style and is intended for command-line or script-based use. DASH is highly modular, and so its various methods can be applied to a diverse range of time periods, spatial regions, proxy networks, and assimilation methods.

DASH contains several key modules. The first module, `gridfile`, is designed to organize and consolidate data sets stored across multiple files and data formats. This module provides a common interface to access data saved across a variety of formats and allows users to manipulate data using metadata values they find meaningful. The second module, `stateVector`, allows users to design and build state vector ensembles with minimal data manipulation. The module allows flexible and modular state vector design, permitting the inclusion of different variables, spatial means, temporal means, and sequences through time. Another module, `ensembleMetadata`, associates the data in a state vector ensemble with user-provided metadata. This allows users to quickly locate values needed to run proxy system models, or needed for post-

processing tasks, such as covariance localization. The `PSM` (proxy-system model) module provides access to numerous proxy system models within the DASH framework. Combined with `ensembleMetadata`, users can efficiently run forward models and align modeled proxy estimates within a state vector ensemble. Finally, the `kalmanFilter`, `particleFilter`, and `optimalSensor` modules implement common DA algorithms in computationally efficient manners. These modules also allow users to quickly select different algorithm parameters, such as covariance localization, or choice of particle filter weights.

In Appendix C, we detail these modules and provide example cases implementing recent scientific studies. Although this dissertation has focused on Common Era DA, the highly modular nature of DASH supports its use over a variety of timescales. Consequently, these examples include cases from both the Common Era and the Last Glacial Maximum. Finally, this appendix discusses DASH in the broader context of paleoclimate data assimilation and describes anticipated future developments.

### 1.3 References

- Abram, N. J., Mulvaney, R., Vimeux, F., Phipps, S. J., Turner, J., and England, M. H.: Evolution of the Southern Annular Mode during the past millennium, *Nature Climate Change*, 4, 564–569, 2014.
- Acevedo, W., Fallah, B., Reich, S., and Cubasch, U.: Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model, *Climate of the Past*, 13, 545–557, 2017.
- Alley, R. B.: Palaeoclimatic insights into future climate challenges, *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361, 1831–1849, 2003.
- Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D’Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B. E., et al.: Last millennium Northern

- Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions, *Quaternary Science Reviews*, 163, 1–22, 2017.
- Annan, J. and Hargreaves, J.: Identification of climatic state with limited proxy data, *Climate of the Past*, 8, 1141–1151, 2012.
- Ault, T., Deser, C., Newman, M., and Emile-Geay, J.: Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium, *Geophysical Research Letters*, 40, 3450–3456, 2013.
- Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., and Meko, D. M.: Assessing the risk of persistent drought using climate model simulations and paleoclimate data, *Journal of Climate*, 27, 7529–7549, 2014.
- Bell, A. R., Osgood, D. E., Cook, B. I., Anchukaitis, K. J., McCarney, G. R., Greene, A. M., Buckley, B. M., and Cook, E. R.: Paleoclimate histories improve access and sustainability in index insurance programs, *Global environmental change*, 23, 774–781, 2013.
- Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, *Climate of the Past*, 8, 963–976, 2012.
- Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., and Otto-Bliesner, B. L.: Pliocene and Eocene provide best analogs for near-future climates, *Proceedings of the National Academy of Sciences*, 115, 13 288–13 293, 2018.
- Cane, M. A., Braconnot, P., Clement, A., Gildor, H., Joussaume, S., Kageyama, M., Khodri, M., Paillard, D., Tett, S., and Zorita, E.: Progress in paleoclimate modeling, *Journal of Climate*, 19, 5031–5057, 2006.
- Coats, S., Smerdon, J., Stevenson, S., Fasullo, J., Otto-Bliesner, B., and Ault, T.: Paleoclimate constraints on the spatiotemporal character of past and future droughts, *Journal of Climate*, 33, 9883–9903, 2020.

- Cook, B. I., Cook, E. R., Anchukaitis, K. J., Seager, R., and Miller, R. L.: Forced and unforced variability of twentieth century North American droughts and pluvials, *Climate dynamics*, 37, 1097–1110, 2011.
- Cort, G. D., Chevalier, M., Burrough, S. L., Chen, C. Y., and Harrison, S. P.: An uncertainty-focused database approach to extract spatiotemporal trends from qualitative and discontinuous lake-status histories, *Quaternary Science Reviews*, 258, 106 870, <https://doi.org/10.1016/j.quascirev.2021.106870>, 2021.
- Crowley, T. J.: Utilization of paleoclimate results to validate projections of a future greenhouse warming, in: *Developments in atmospheric science*, vol. 19, pp. 35–45, Elsevier, 1991.
- Dätwyler, C., Neukom, R., Abram, N. J., Gallant, A. J., Grosjean, M., Jacques-Coper, M., Karoly, D. J., and Villalba, R.: Teleconnection stationarity, variability and trends of the Southern Annular Mode (SAM) during the last millennium, *Climate Dynamics*, 51, 2321–2339, 2018.
- Dee, S. G., Steiger, N. J., Emile-Geay, J., and Hakim, G. J.: On the utility of proxy system models for estimating climate states over the Common Era, *Journal of Advances in Modeling Earth Systems*, 8, 1164–1179, 2016.
- Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, *Climate dynamics*, 38, 527–546, 2012.
- Dirren, S. and Hakim, G. J.: Toward the assimilation of time-averaged observations, *Geophysical research letters*, 32, 2005a.
- Dirren, S. and Hakim, G. J.: Toward the assimilation of time-averaged observations, *Geophysical research letters*, 32, 2005b.
- Dubinkina, S., Goosse, H., Sallaz-Damaz, Y., Crespin, E., and Crucifix, M.: Testing a particle filter to reconstruct climate changes over the past centuries, *International Journal of Bifurcation and Chaos*, 21, 3611–3618, 2011.

- Esper, J., Frank, D. C., Wilson, R. J., and Briffa, K. R.: Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophysical Research Letters*, 32, 2005.
- Esper, J., St George, S., Anchukaitis, K., D'Arrigo, R., Ljungqvist, F. C., Luterbacher, J., Schneider, L., Stoffel, M., Wilson, R., and Buntgen, U.: Large-scale, millennial-length temperature reconstructions from tree-rings, *Dendrochronologia*, 50, 81–90, <https://doi.org/10.1016/j.dendro.2018.06.001>, 2018.
- Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343–367, 2003.
- Fang, S.-w., Khodri, M., Timmreck, C., Zanchettin, D., and Jungclaus, J.: Disentangling internal and external contributions to Atlantic multidecadal variability over the past millennium, *Geophysical Research Letters*, 48, e2021GL095 990, 2021.
- Fernández-Donado, L., González-Rouco, J., Raible, C., Ammann, C., Barriopedro, D., García-Bustamante, E., Jungclaus, J. H., Lorenz, S., Luterbacher, J., Phipps, S. J., et al.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, *Climate of the Past*, 9, 393–421, 2013.
- Frank, D., Esper, J., Zorita, E., and Wilson, R.: A noodle, hockey stick, and spaghetti plate: a perspective on high-resolution paleoclimatology, *Wiley Interdisciplinary Reviews: Climate Change*, 1, 507–516, 2010.
- Franke, J., Valler, V., Brönnimann, S., Neukom, R., and Jaume-Santero, F.: The importance of input data quality and quantity in climate field reconstructions – results from the assimilation of various tree-ring collections, *Climate of the Past*, 16, 1061–1074, <https://doi.org/10.5194/cp-16-1061-2020>, 2020.
- Gallant, A. J., Phipps, S. J., Karoly, D. J., Mullan, A. B., and Lorrey, A. M.: Nonstationary Australasian teleconnections and implications for paleoclimate reconstructions, *Journal of Climate*, 26, 8827–8849, 2013.

- Gillett, N. P., Kell, T. D., and Jones, P.: Regional climate impacts of the Southern Annular Mode, *Geophysical Research Letters*, 33, 2006.
- Gong, D. and Wang, S.: Definition of Antarctic oscillation index, *Geophysical Research Letters*, 26, 459–462, 1999.
- Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M. E., Renssen, H., Sallaz-Damaz, Y., and Shindell, D.: The role of forcing and internal dynamics in explaining the “Medieval Climate Anomaly”, *Climate dynamics*, 39, 2847–2866, 2012a.
- Goosse, H., Guiot, J., Mann, M. E., Dubinkina, S., and Sallaz-Damaz, Y.: The Medieval Climate Anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation, *Global and Planetary Change*, 84, 35–47, 2012b.
- Guillet, S., Corona, C., Stoffel, M., Khodri, M., Lavigne, F., Ortega, P., Eckert, N., Sienou, P. D., Daux, V., Churakova, O. v., et al.: Climate response to the Samalas volcanic eruption in 1257 revealed by proxy records, *Nature Geoscience*, 10, 123–128, 2017.
- Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., Gong, D., Kaufman, D. S., Nnamchi, H. C., Quaas, J., Rivera, J. A., Sathyendranath, S., Smith, S. L., Trewin, B., von Shuckmann, K., and Vose, R. S.: Changing State of the Climate System, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., chap. 2, Cambridge University Press, 2021.
- Gupta, A. S. and England, M. H.: Coupled ocean–atmosphere–ice response to variations in the southern annular mode, *Journal of Climate*, 19, 4457–4486, 2006.
- Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The Last Millennium Climate Reanalysis project: Framework

- and first results, *Journal of Geophysical Research: Atmospheres*, 121, 6745–6764, 2016.
- Hargreaves, J. and Annan, J.: On the importance of paleoclimate modelling for improving predictions of future climate change, *Climate of the Past*, 5, 803–814, 2009.
- Hargreaves, J., Abe-Ouchi, A., and Annan, J.: Linking glacial and future climates through an ensemble of GCM simulations, *Climate of the Past*, 3, 77–87, 2007.
- Hendon, H. H., Thompson, D. W., and Wheeler, M. C.: Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode, *Journal of Climate*, 20, 2452–2467, 2007.
- Hessl, A., Allen, K. J., Vance, T., Abram, N. J., and Saunders, K. M.: Reconstructions of the southern annular mode (SAM) during the last millennium, *Progress in Physical Geography*, 41, 834–849, 2017.
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S., Danabasoglu, G., Edwards, J., et al.: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability, *Bulletin of the American Meteorological Society*, 96, 1333–1349, 2015.
- King, J., Anchukaitis, K., Allen, K., Vance, T., and Hessl, A.: Trends and variability in the Southern Annular Mode over the Common Era, in review, 2022.
- King, J. M., Anchukaitis, K. J., Tierney, J. E., Hakim, G. J., Emile-Geay, J., Zhu, F., and Wilson, R.: A data assimilation approach to last millennium temperature field reconstruction using a limited high-sensitivity proxy network, *Journal of Climate*, pp. 1–64, 2021.
- Kwok, R. and Comiso, J. C.: Spatial patterns of variability in Antarctic surface temperature: Connections to the Southern Hemisphere Annular Mode and the Southern Oscillation, *Geophysical Research Letters*, 29, 50–1, 2002.

- Liu, H., Liu, Z., and Lu, F.: A Systematic Comparison of Particle Filter and EnKF in Assimilating Time-Averaged Observations, *Journal of Geophysical Research: Atmospheres*, 122, 13–155, 2017.
- Mairesse, A., Goosse, H., Mathiot, P., Wanner, H., and Dubinkina, S.: Investigating the consistency between proxy-based reconstructions and climate models using data assimilation: a mid-Holocene case study, *Climate of the Past*, 9, 2741–2757, 2013.
- Mann, M. E. and Rutherford, S.: Climate reconstruction using ‘Pseudoproxies’, *Geophysical Research Letters*, 29, 139–1, 2002.
- Marshall, G. J.: Trends in the Southern Annular Mode from observations and reanalyses, *Journal of Climate*, 16, 4134–4143, 2003.
- Marshall, G. J., Orr, A., Van Lipzig, N. P., and King, J. C.: The impact of a changing Southern Hemisphere Annular Mode on Antarctic Peninsula summer temperatures, *Journal of Climate*, 19, 5388–5404, 2006.
- Matsikaris, A., Widmann, M., and Jungclauss, J. H.: On-line and off-line data assimilation in palaeoclimatology: a case study, *Climate of the Past*, 11, 81–93, 2015.
- Morales, M. S., Cook, E. R., Barichivich, J., Christie, D. A., Villalba, R., LeQuesne, C., Srur, A. M., Ferrero, M. E., González-Reyes, Á., Couvreur, F., et al.: Six hundred years of South American tree rings reveal an increase in severe hydroclimatic events since mid-20th century, *Proceedings of the National Academy of Sciences*, 117, 16 816–16 823, 2020.
- Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufman, D. S., Lücke, L., Rehfeld, K., et al.: Consistent multi-decadal variability in global temperature reconstructions and simulations over the Common Era, *Nature geoscience*, 12, 643, 2019a.
- Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P.: No evidence for globally coherent warm and cold periods over the preindustrial Common Era, *Nature*, 571, 550–554, 2019b.

- Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., and Kosro, P. M.: Assimilation of surface velocity data into a primitive equation coastal ocean model, *Journal of Geophysical Research: Oceans*, 107, <https://doi.org/10.1029/2000JC000511>, 2002.
- Osman, M. B., Tierney, J. E., Zhu, J., Tardif, R., Hakim, G. J., King, J., and Poulsen, C. J.: Globally resolved surface temperatures since the Last Glacial Maximum, *Nature*, 599, 239–244, 2021.
- Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G.: Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model, *Bulletin of the American Meteorological Society*, 97, 735–754, 2016.
- Overpeck, J. T., Otto-Bliesner, B. L., Miller, G. H., Muhs, D. R., Alley, R. B., and Kiehl, J. T.: Paleoclimatic evidence for future ice-sheet instability and rapid sea-level rise, *Science*, 311, 1747–1750, 2006.
- PAGES 2k-PMIP3 group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, *Climate of the Past*, 11, 1673–1699, <https://doi.org/10.5194/cp-11-1673-2015>, 2015.
- PAGES2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, *Scientific Data*, 4, 2017.
- Palmer, J. G., Cook, E. R., Turney, C. S., Allen, K., Fenwick, P., Cook, B. I., O'Donnell, A., Lough, J., Grierson, P., and Baker, P.: Drought variability in the eastern Australia and New Zealand summer drought atlas (ANZDA, CE 1500–2012) modulated by the Interdecadal Pacific Oscillation, *Environmental Research Letters*, 10, 124 002, 2015.
- Parsons, L. A., Amrhein, D. E., Sanchez, S. C., Tardif, R., Brennan, M. K., and Hakim, G. J.: Do Multi-Model Ensembles Improve Reconstruction Skill in Paleoclimate Data Assimilation?, *Earth and Space Science*, 8, e2020EA001 467, 2021.
- Perkins, W. A. and Hakim, G. J.: Reconstructing paleoclimate fields using online data assimilation with a linear inverse model, *Climate of the Past*, 13, 421–436, 2017.

- Rice, J. L., Woodhouse, C. A., and Lukas, J. J.: Science and Decision Making: Water Management and Tree-Ring Data in the Western United States, *JAWRA Journal of the American Water Resources Association*, 45, 1248–1259, 2009.
- Rogers, J. C. and Van Loon, H.: Spatial variability of sea level pressure and 500 mb height anomalies over the Southern Hemisphere, *Monthly Weather Review*, 110, 1375–1392, 1982.
- Schmidt, G., Annan, J., Bartlein, P., Cook, B., Guilyardi, E., Hargreaves, J., Harrison, S., Kageyama, M., LeGrande, A., Konecky, B., et al.: Using paleo-climate comparisons to constrain future projections in CMIP5., *Climate of the Past Discussions*, 9, 2013.
- Schmidt, G. A.: Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change, *Journal of Quaternary Science*, 25, 79–87, 2010.
- Silvestri, G. and Vera, C.: Nonstationary impacts of the southern annular mode on Southern Hemisphere climate, *Journal of Climate*, 22, 6142–6148, 2009.
- Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudo-proxy experiments, *Wiley Interdisciplinary Reviews: Climate Change*, 3, 63–77, 2012.
- Snyder, C. W.: The value of paleoclimate research in our changing climate, *Climatic Change*, 100, 407–418, 2010.
- Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of time-averaged pseudoproxies for climate reconstruction, *Journal of Climate*, 27, 426–441, 2014.
- Steiger, N. J., Steig, E. J., Dee, S. G., Roe, G. H., and Hakim, G. J.: Climate reconstruction using data assimilation of water isotope ratios from ice cores, *Journal of Geophysical Research: Atmospheres*, 122, 1545–1568, 2017.
- Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, *Scientific Data*, 5, <https://doi.org/10.1086/sdata.2018.86>, 2018.

- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling., *Climate of the Past*, 15, 1251–1273, 2019.
- Thompson, D. W. and Solomon, S.: Interpretation of recent Southern Hemisphere climate change, *Science*, 296, 895–899, 2002.
- Thompson, D. W. and Wallace, J. M.: Annular modes in the extratropical circulation. Part I: Month-to-month variability, *Journal of Climate*, 13, 1000–1016, 2000.
- Tierney, J. E., Poulsen, C. J., Montañez, I. P., Bhattacharya, T., Feng, R., Ford, H. L., Hönisch, B., Inglis, G. N., Petersen, S. V., Sago, N., et al.: Past climates inform our future, *Science*, 370, eaay3701, 2020a.
- Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., and Poulsen, C. J.: Glacial cooling and climate sensitivity revisited, *Nature*, 584, 569–573, 2020b.
- Valler, V., Franke, J., and Brönnimann, S.: Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation, *Climate of the Past*, 15, 1427–1441, <https://doi.org/10.5194/cp-15-1427-2019>, 2019.
- Van Lipzig, N. P., Marshall, G. J., Orr, A., and King, J. C.: The relationship between the Southern Hemisphere Annular Mode and Antarctic Peninsula summer temperatures: Analysis of a high-resolution model climatology, *Journal of Climate*, 21, 1649–1668, 2008.
- Villalba, R., Lara, A., Masiokas, M. H., Urrutia, R., Luckman, B. H., Marshall, G. J., Mundo, I. A., Christie, D. A., Cook, E. R., Neukom, R., et al.: Unusual Southern Hemisphere tree growth patterns induced by changes in the Southern Annular Mode, *Nature Geoscience*, 5, 793–798, 2012.

- Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., and Rajaratnam, B.: Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation, *Geophysical Research Letters*, 42, 7162–7170, 2015.
- Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D’arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., et al.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, *Quaternary Science Reviews*, 134, 1–18, 2016.
- Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., and Anchukaitis, K. J.: Resolving the differences in the simulated and reconstructed temperature response to volcanism, *Geophysical Research Letters*, 47, e2019GL086908, 2020.
- Zhu, F., Emile-Geay, J., Anchukaitis, K. J., Hakim, G. J., Wittenberg, A. T., Morales, M. S., Toohey, M., and King, J.: A re-appraisal of the ENSO response to volcanism with paleoclimate data assimilation, *Nature communications*, 13, 1–9, 2022.
- Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Gettelman, A., Bacmeister, J. T., Neale, R. B., Poulsen, C. J., Shaw, J. K., McGraw, Z. S., and Kay, J. E.: LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2, *Journal of Advances in Modeling Earth Systems*, p. e2021MS002776, 2021a.
- Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Poulsen, C. J., Tierney, J. E., Lofverstrom, M., and DiNezio, P.: Assessment of equilibrium climate sensitivity of the Community Earth System Model version 2 through simulation of the Last Glacial Maximum, *Geophysical Research Letters*, 48, e2020GL091220, 2021b.
- Zorita, E., González-Rouco, F., and Legutke, S.: Testing the approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model, *Journal of Climate*, 16, 1378–1390, 2003.

## APPENDIX A

A data assimilation approach to Last Millennium temperature field reconstruction using a limited high-sensitivity proxy network

King, J., Anchukaitis, K., Tierney, J., Hakim, G., Emile-Geay, J., Zhu, F., Wilson, R. (2021) A data assimilation approach to Last Millennium temperature field reconstruction using a limited high-sensitivity proxy network. *Journal of Climate*, 34(17), 7091-7111.

doi:JCLI-D-20-0661.1

## **A data assimilation approach to Last Millennium temperature field reconstruction using a limited high-sensitivity proxy network**

Jonathan King<sup>1,2</sup>, Kevin Anchukaitis<sup>2,3,1</sup>, Jessica Tierney<sup>1</sup>, Gregory Hakim<sup>4</sup>, Julien Emile-Geay<sup>5</sup>, Feng Zhu<sup>5</sup>, Rob Wilson<sup>6</sup>

<sup>1</sup> Department of Geosciences, University of Arizona, Tucson, Arizona, USA

<sup>2</sup> Laboratory of Tree-Ring Research, University of Arizona, Tucson, Arizona, USA

<sup>3</sup> School of Geography, Development, and Environment, University of Arizona, Tucson, Arizona, USA

<sup>4</sup> Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA

<sup>5</sup> Department of Earth Sciences, University of Southern California, Los Angeles, California, USA

<sup>6</sup> School of Earth and Environmental Sciences, University of St Andrews, St Andrews, UK

### **A.1 Abstract**

We use the Northern Hemisphere Tree-Ring Network Development (NTREND) tree-ring database to examine the effects of using a small, highly-sensitive proxy network for paleotemperature data assimilation over the last millennium. We first evaluate our methods using pseudo-proxy experiments. These indicate that spatial assimilations using this network are skillful in the extratropical Northern Hemisphere and improve on previous NTREND reconstructions based on Point-by-Point regression. We also find our method is sensitive to climate model biases when the number of sites becomes small. Based on these experiments, we then assimilate the real NTREND network. To quantify model prior uncertainty, we produce 10 separate reconstructions, each assimilating a different climate model. These reconstructions are most dissimilar prior to 1100 CE, when the network becomes sparse, but show greater consistency as the network grows. Temporal variability is also underestimated before 1100 CE. Our assimilation method produces spa-

tial uncertainty estimates and these identify treeline North America and eastern Siberia as regions that would most benefit from development of new millennial-length temperature-sensitive tree-ring records. We compare our multi-model mean reconstruction to five existing paleo-temperature products to examine the range of reconstructed responses to radiative forcing. We find substantial differences in the spatial patterns and magnitudes of reconstructed responses to volcanic eruptions and in the transition between the Medieval epoch and Little Ice Age. These extant uncertainties call for the development of a paleoclimate reconstruction intercomparison framework for systematically examining the consequences of proxy network composition and reconstruction methodology and for continued expansion of tree-ring proxy networks.

## **A.2 Introduction**

Past variations in surface temperatures can be used to investigate a number of key characteristics of the Earth's climate system, including the response to radiative forcing, the regional effects of such forcings, and the role of internal modes of coupled ocean-atmosphere variability (Hegerl et al., 1997; Stott and Tett, 1998; Delworth and Mann, 2000; Meehl et al., 2004; Lean and Rind, 2008; Stott and Jones, 2009; Stott et al., 2010; Solomon et al., 2011; Phipps et al., 2013; Hegerl and Stott, 2014; Kaufman, 2014; Guillet et al., 2017; Neukom et al., 2019; Zhu et al., 2020). Paleoclimate temperature reconstructions using natural archives like tree-rings are particularly useful because they extend the short instrumental record to centennial and longer timescales. These provide an opportunity to characterize the patterns and magnitude of forced climate response and internal variability (Hegerl et al., 2003, 2007; Schurer et al., 2013; Masson-Delmotte et al., 2013). Climate field reconstructions (CFRs) can additionally capture the spatial fingerprints of large-scale temperature anomalies caused by radiative forcing and ocean-atmosphere dynamics (Mann et al., 1998; Evans et al., 2001; Seager et al., 2007; Cook et al., 2010a,b; Phipps et al., 2013; Anchukaitis and McKay, 2015; Goosse, 2017). CFRs have been developed using a number of methods (Tingley et al., 2012; Smerdon and Pollack, 2016) including point-by-point methods (Cook et al., 1999, 2010a,b; Anchukaitis et al., 2017),

variants of regularized expectation maximization (RegEM; Schneider, 2001; Rutherford et al., 2003; Mann et al., 2009; Smerdon et al., 2011; Guillot et al., 2015), and reduced space approaches (Fritts, 1991; Cook et al., 1994; Mann et al., 1998; Evans et al., 2002; Gill et al., 2016).

Recently, data assimilation (DA) has emerged as a promising CFR technique (e.g. Widmann et al., 2010; Bhend et al., 2012; Goosse et al., 2012; Steiger et al., 2014; Hakim et al., 2016; Matsikaris et al., 2015; Okazaki and Yoshimura, 2017; Steiger et al., 2018; Franke et al., 2020). Assimilation methods integrate the climate signals recorded in paleoclimate proxies with dynamical constraints provided by climate models to produce spatially continuous climate field reconstructions and associated uncertainty estimates. There are several existing paleoclimate DA paradigms, including pattern nudging / forcing singular vectors (Van der Schrier and Barkmeijer, 2005), particle filters (Goosse et al., 2012; Dubinkina and Goosse, 2013; Matsikaris et al., 2015), and ensemble Kalman filters (Bhend et al., 2012; Steiger et al., 2014; Hakim et al., 2016; Dee et al., 2016; Perkins and Hakim, 2017; Steiger et al., 2018; Tardif et al., 2019; Franke et al., 2020). Here, we focus on the ensemble Kalman filter (EnKF) approach (Steiger et al., 2014; Hakim et al., 2016), which has been shown to perform well compared to other DA methods in a paleoclimate context (Liu et al., 2017). EnKF methods update an ensemble of climate states to more closely match paleoclimate proxy records. These climate states are produced using one of two approaches: the “online” method, in which the ensemble is generated by a set of transient model simulations that propagate updates forward through time (e.g. Perkins and Hakim, 2017); and the “offline” (or “no-cycling”) method (Oke et al., 2002; Evensen, 2003), in which ensembles are constructed from pre-existing climate model output (e.g. Bhend et al., 2012; Annan and Hargreaves, 2012; Steiger et al., 2014; Hakim et al., 2016; Valler et al., 2019; Tardif et al., 2019; Franke et al., 2020). We focus here on the offline approach, which has been shown to perform favorably to online methods in paleoclimate contexts with reduced computational costs (Matsikaris et al., 2015; Acevedo et al., 2017). A key requirement of EnKF methods is the ability to estimate equivalent proxy values from climate model output. This is achieved through the use of forward models that translate climate state variables, like surface temperature, into proxy values, like tree-ring

width (TRW) or maximum latewood density (MXD). These forward models can range in complexity from a simple linear relationship to more detailed Proxy Systems Models (PSMs) incorporating the physical processes that transform climate signals to proxy records (Evans et al., 2013). The use of forward models helps separate data and process level models in the data assimilation framework (Goosse, 2016).

An important decision in any assimilation is the selection of the proxy network. Ultimately, this choice must balance spatiotemporal coverage with sensitivity to the reconstructed field and associated proxy uncertainties (Esper et al., 2005; Frank et al., 2010; Wang et al., 2015; Wilson et al., 2016; Anchukaitis et al., 2017; Esper et al., 2018; Franke et al., 2020; Cort et al., 2021). In general, large networks maximize coverage, but their size often results from the inclusion of proxy records with comparatively weak, complex, seasonally varying, or multivariate sensitivity to reconstructed variables. By contrast, smaller curated networks consisting of well-understood and strongly-sensitive proxies provide a higher ratio of signal to noise at the cost of reduced coverage (Frank et al., 2010). An additional consideration concerns the implementation of forward models: highly sensitive networks with a known climate response and seasonal window facilitate physically realistic forward models, potentially improving assimilation skill. Given the complexity of these trade-offs, network selection is not necessarily intuitive. Noisy proxies that covary poorly with climate fields are down-weighted by the Kalman filter algorithm; if this down-weighting renders the effects of climate-insensitive proxies negligible on a reconstruction, then a large network incorporating many proxies might appear preferable. However, work by Franke et al. (2020) indicates that EnKF temperature reconstructions using large proxy networks do not correlate with target temperatures as well as reconstructions produced using smaller, more sensitive networks. This result is supported by Tardif et al. (2019), who found that additional screening of proxy records for temperature sensitivity in an assimilation framework improved their ability to reconstruct salient pre-industrial climate features, such as cooling during the Little Ice Age. The importance of proxy sensitivity is further highlighted by Steiger and Smerdon (2017) who note that skillful hydroclimate DA requires proxies sensitive to the target reconstruction field.

Curated temperature sensitive proxy networks for data assimilation include the PAGES2k (PAGES2k Consortium, 2013, 2017) and NTREND networks (Wilson et al., 2016; Anchukaitis et al., 2017). The PAGES2k network has been commonly used in paleo-DA applications (Hakim et al., 2016; Dee et al., 2016; Okazaki and Yoshimura, 2017; Perkins and Hakim, 2017; Tardif et al., 2019; Neukom et al., 2019) and consists of proxy records identified as temperature-sensitive and meeting minimum temporal coverage and age model precision criteria during the Common Era (PAGES2k Consortium, 2017). DA reconstructions using this network may implement additional proxy screening but usually incorporate several hundred proxy records. The NTREND network has stricter requirements for inclusion: it consists of 54 published tree-ring chronologies selected by dendroclimatologists for demonstrating an established and reasonable biophysical association with local seasonal temperatures (Wilson et al., 2016). Franke et al. (2020) proposed that the additional coverage of the PAGES2k network is preferable to the increased sensitivity of the smaller NTREND network for global and hemisphere-scale temperature reconstructions but found the NTREND network provided the best reconstruction in the extratropical Northern Hemisphere. To produce a maximally skillful reconstruction for this region, we focus on assimilating the NTREND network but acknowledge that this choice is accompanied by a reduced spatial extent.

Before performing an assimilation, we seek to understand the advantages and trade-offs of offline EnKF related to both the proxy data and climate model priors. We implement these sensitivity tests using pseudo-proxy experiments (Mann and Rutherford, 2002; Zorita et al., 2003; Smerdon, 2012), which allow us to test the DA method’s ability to reconstruct known climate fields within a controlled setting. Here, we note the importance of model selection in DA pseudo-proxy experiments and distinguish between “perfect-model” and “biased-model” experimental designs. In a perfect-model experiment, the same model is used to generate the target field and as the model prior. Such designs are common in DA analyses (Annan and Hargreaves, 2012; Steiger et al., 2014; Okazaki and Yoshimura, 2017; Acevedo et al., 2017; Zhu et al., 2020), where they are powerful tools for testing sensitivity to variables like proxy noise, network distribution, and calibration intervals. Biased-model paradigms use different climate models to gener-

ate target fields and assimilated model priors and can help examine the effects of biases in a model prior's mean state and spatial covariance. Dee et al. (2016) found model biases a potentially major source of error in paleo-EnKF reconstructions, so we employ both perfect and biased-model experiments in our investigations.

In this study, we begin by first evaluating the sensitivity of our DA method to proxy noise, network attrition, and climate model biases in a suite of pseudo-proxy experiments. We also use the pseudo-proxy framework to compare the skill of our DA method to point-by-point regression (PPR), the technique used for the original NTREND temperature field reconstruction (Anchukaitis et al., 2017). We then assimilate the real NTREND tree-ring network to reconstruct mean May through August (MMJA) temperature anomalies. We produce an ensemble of real reconstructions by assimilating NTREND with output from multiple climate models in the Coupled Modeling Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2012) and the Community Earth System Model (CESM) Last Millennium Ensemble (LME; Otto-Bliesner et al., 2016). We quantify the skill of the DA reconstructions using spatial temperature anomaly fields, mean Northern Hemisphere extratropical (30°N–90°N) May through August time series, and withheld proxy data. Finally, we examine the climate response of the ensemble-mean reconstruction to radiative forcings and compare these responses against existing temperature field reconstructions.

### **A.3 Methods**

#### **A.3.1 Proxy Network**

The NTREND network is a curated set of 54 published annual resolution tree-ring based summer-temperature proxy records selected by dendroclimatologists to maximize sensitivity to boreal summer temperatures while minimizing the response to other climate variables (Figure A.1; Wilson et al., 2016; Anchukaitis et al., 2017). Although tree growth at the NTREND sites is primarily limited by summer growing temperatures, the optimal summer season varies between sites. Wilson et al. (2016) determined the season of highest temperature sensitivity for each site and identified mean MJJA temperatures anomalies as the optimal reconstruction target for the network as a whole. The network

only includes sites between 40°N and 75°N as lower latitude trees tend to exhibit sensitivity to multiple climate influences, especially moisture limitations. Each record is derived from ring-width measurements (TRW), maximum latewood density (MXD; Schweingruber et al., 1978), or a mixture of TRW, MXD, and blue intensity (BI; McCarroll et al., 2002; Björklund et al., 2014; Rydval et al., 2014; Wilson et al., 2019). The network extends from 750 - 2011 CE, with maximum coverage over the period from 1710-1988 CE. Spatial coverage is greater over Eurasia (39 sites) than North America (15 sites), with a distinct spatial imbalance prior to 1000 CE (20 vs. 3). We end all reconstructions in 1988 CE as network attrition limits the utility of assimilated NTREND reconstructions after this point (Anchukaitis et al., 2017).

### A.3.2 Data Assimilation

Our data assimilation method uses an ensemble Kalman filter (EnKF) (Evensen, 1994; Steiger et al., 2014)

$$\mathbf{X}_a = \mathbf{X}_p + \mathbf{K}(\mathbf{Y} - \mathbf{Y}_e) \quad (\text{A.1})$$

to update an initial ensemble of climate states ( $\mathbf{X}_p$ ) given proxy data ( $\mathbf{Y}$ ) and model estimates of the proxy data ( $\mathbf{Y}_e$ ). These data are combined via the Kalman Gain ( $\mathbf{K}$ ; detailed in Section A.8.1) to produce an updated ensemble ( $\mathbf{X}_a$ ) in each reconstructed annual time step. We use an EnKF variant known as the ensemble square root Kalman filter (EnSRF; Andrews, 1968), with an “offline” (or “no-cycling”) approach (Oke et al., 2002; Evensen, 2003). The complete details of our approach are given in section A.8.1 and described in Steiger et al. (2014) and Hakim et al. (2016). The Kalman Filter can be expressed as a recursive Bayesian filter (Chen et al., 2003; Wikle and Berliner, 2007), wherein new information ( $\mathbf{Y}$ ) updates estimates of state parameters ( $\mathbf{X}$ ). Hence, we will often refer to  $\mathbf{X}_p$  as the model prior, and the updated ensemble  $\mathbf{X}_a$  as the model posterior.

We implement a covariance localization scheme, which limits the influence of proxies outside of a specified radius. Localization was originally developed to limit spurious covariance arising from sampling noise in small ensembles of  $m \leq 50$  (Houtekamer

and Mitchell, 2001). Our offline approach enables the use of much larger ensembles ( $m > 1000$ ), but we note that spurious covariances may still arise from biases in a climate model’s covariance structure. Consequently, localization may improve the quality of assimilated paleoclimate reconstructions even for large prior ensembles. The localization radius is an important free parameter in this method and must be assessed independently for different model priors, reconstruction targets, and proxy networks (Tables A.2, A.6). The process used to select localization radii for these experiments is detailed in Section A.8.1.

To generate model estimates of the proxy values, we follow the methodology of Tardif et al. (2019) and use linear univariate forward models trained on the mean temperature of each site’s optimal growing season (Wilson et al., 2016), such that:

$$\mathbf{y}_{\mathbf{e}_j} = \alpha_j + \beta_j \mathbf{T}_j. \quad (\text{A.2})$$

Here,  $\mathbf{T}_j$  is a vector of mean growing-season temperature anomalies extracted from the prior. The coefficients  $\alpha_j$  and  $\beta_j$  are determined by regressing assimilated observations ( $\hat{\mathbf{y}}_j$ ) against mean growing-season temperature anomalies from the closest grid cell of the target field. We emphasize that these target fields vary by application. For pseudo-proxy experiments, the target field is a specific model realization, whereas the real assimilation uses CRU-TS 4.01 (Harris et al., 2014). Regardless of the target, we perform each regression over the years in which the real NTREND records overlap data from the closest land grid cell in CRU-TS 4.01; this ensures that both pseudo-proxy and real reconstructions use regressions with the same temporal span. The variance of each record’s regression residuals is used as the observation uncertainty ( $R_{jj}$ ) in the Kalman Filter (Section A.8.1). This uncertainty ranges from 0.23 to 1.34 proxy units over the network.

We construct prior ensembles using output from the past1000 and historical experiments of the Coupled Modeling Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2012) as well as the Last Millennium Ensemble (LME; Otto-Bliesner et al., 2016). For a given assimilation, we use values from a single climate model and designate each year of available output as a unique ensemble member. We use static model priors, whereby the same prior is used for each reconstructed time step. This scheme is justified by the lim-

ited forecast skill of climate models beyond the annual reconstruction timescale (Bhend et al., 2012) and is common in paleo-DA applications (e.g. Steiger et al., 2014; Dee et al., 2016; Tardif et al., 2019). A summary of the model ensembles is given in Table A.1. The past1000 CMIP5 data for each model are from the ensemble member designated *r1i1p1*, and LME output was selected from full-forcing run 2. We assimilate temperature anomalies relative to the 1951-1980 CE mean; this helps avoid the effects of climate model mean state biases, but we note that model covariance biases are unaffected. In all reconstructions, we update the mean May through August (MJJJA) temperature anomaly field, rather than individual months. We assess the skill of each assimilation by comparing the Pearson’s correlation coefficients, root mean square errors (RMSEs), mean biases, and standard deviation ratios.

### A.3.3 Pseudo-proxy Reconstructions

Before assimilating the real NTREND network, we first examine the skill of our DA method in a pseudo-proxy framework (Smerdon, 2012). This approach allows us to test the method’s ability to reconstruct known climate field targets within a controlled setting. Here, we specify the target fields as surface temperatures from the years 850-2005 CE from either the Last Millennium Ensemble full-forcing run 2 (CESM; Otto-Bliesner et al., 2016), or from the combined last millennium and historical runs of the Max Planck Institute for Meteorology Earth System Model (MPI; Marsland et al., 2003; Stevens et al., 2013). While this experimental design is intentionally tractable, we caution that the observed spatial patterns of skill will depend on the specific models used (Smerdon et al., 2011). Here, we are interested in examining the sensitivity of EnSRF to the proxy network and climate model prior, so we systematically explore the effects of noisy proxy records, network attrition, and biased climate models on DA performance. To examine the effects of model covariance biases, we test each combination of target field and model prior for LME and MPI, which allows us to alternate between perfect-model and biased-model experimental designs.

After selecting a target field, we generate pseudo-proxies using:

$$\hat{y}_j = a_j + b_j \mathbf{T}_j^{\text{target}} + \epsilon_j \quad (\text{A.3})$$

where  $\hat{y}_j$  is the  $j^{\text{th}}$  pseudo-proxy record and  $\mathbf{T}_j^{\text{target}}$  is the vector of mean growing season temperature anomalies from the grid cell closest to the proxy site in the target climate field. The coefficients  $a_j$  and  $b_j$  are the intercept and slope obtained by regressing the real NTREND network against mean growing-season temperature anomalies from the nearest land cells in CRU-TS 4.01; in this way, the pseudo-proxies mimic the temperature response of the real NTREND network for at least the instrumental period.

We examine the effects of proxy noise by selectively neglecting or adding Gaussian white noise to the pseudo-proxies, such that:

$$\epsilon_j \sim \begin{cases} 0, & \text{Perfect} \\ \mathcal{N}(0, R_{jj}), & \text{Noisy} \end{cases} \quad (\text{A.4})$$

Here,  $R_{jj}$  is the proxy-uncertainty weight for the  $j^{\text{th}}$  NTREND record and is the variance of the NTREND-CRU regression residuals. When testing noisy proxies, we perform 101 assimilations using different noise matrices and report the median skill metrics. Here, we use white noise because it allows us to directly tune the  $R_{jj}$  weight in the Kalman Filter. The median signal-to-noise ratio is 0.80 for the CESM pseudo-proxies and 0.85 for the MPI pseudo-proxies, which is consistent with values found in other pseudo-proxy experiments (Smerdon, 2012). In each test, we examine the effects of network attrition by first assimilating the full set of pseudo-proxies over the entire period and then comparing this to an assimilation where the pseudo-proxies are subjected to the same temporal attrition as the real NTREND network.

After generating pseudo-proxies for a given experiment, we generate pseudo-proxy estimates by applying equation A.2 to the prior ensemble. The coefficients  $\alpha_j$  and  $\beta_j$  are determined by regressing the pseudo-proxies against the target field. Note that pseudo-proxy noise and sampling errors will affect the statistics obtained from these regressions, so  $\alpha_j$  and  $\beta_j$  are estimates of the coefficients  $a_j$  and  $b_j$  used to generate the pseudo-proxies. This mimics how noise and sampling errors can introduce errors into forward

models calibrated on real NTREND data. Once we obtain pseudo-proxy estimates, we then determine an optimal localization radius (Section A.8.1, Table A.2).

A key feature of pseudo-proxy experiments is that the target reconstruction is known. Consequently, we can assess skill directly against the correct answer. Here, we examine pseudo-proxy reconstruction skill using mean Northern Hemisphere extratropical (30°N–90°N) MJJA temperature time series, and spatial grid point time series over the full reconstruction period (850 CE to 1988 CE).

We compare the most realistic (biased-model, noisy-proxy, temporal-attrition) pseudo-proxy DA reconstructions to analogous reconstructions generated using point-by-point regression (PPR). PPR is a “region of interest” CFR technique that iteratively calculates a nested multivariate principal components regression model between predictor network and each point in the target field (Cook et al., 1999). The method was motivated by the premise that proxies near a reconstructed grid point are more likely to reflect climate at that site. Consequently, PPR uses a strict search radius to select proxy predictor series for each grid point reconstruction. The method was first used for drought reconstructions (Cook et al., 1999, 2010a,b) and later adapted for continental temperature anomalies (Cook et al., 2013). Anchukaitis et al. (2017) used the method to reconstruct hemispheric temperature anomalies, and we follow their implementation in this study.

In brief, given a target of gridded climate observation, the method first identifies proxy sites within 1000 km of each grid point centroid. If no proxy records are found within 1000 km, the search radius is expanded in 500 km increments to a maximum of 2000 km until proxy sites are found within the radius. All proxy sites found within the search radius are then used as predictor sites for that grid point. If no predictors are found within 2000 km, then no reconstruction is performed for the grid. These radii are based on decorrelation decay lengths in the observational temperature field from Cowtan and Way (2014). A multivariate regression model is then calibrated against the MJJA temperature values of the target field (Cowtan and Way, 2014) for each grid point over the period 1945 to 1988 CE, and the reconstructions are validated using withheld temperature data for the period 1901 to 1944 CE. As the number of records declines back through time, the regression model is recalibrated and validated for each change in network size and

scaled to match the mean and variance of the predictand during their overlapping time period (Meko, 1997; Cook et al., 1999). For a given grid point, temperature anomalies are obtained for all years in which at least one predictor record remains within the initial search radius. Following Anchukaitis et al. (2017), we then screen the final reconstructed field in each time step to only include grid cells where the reduction of error (RE; Cook et al. (1994)) statistic is greater than zero. We use this screened field here as the final PPR MJJA temperature reconstruction.

### **A.3.4 Real NTREND Reconstruction**

We next assimilate the real NTREND network. To examine the effects of prior selection, we produce 10 real DA reconstructions each using a different climate model to generate the prior (Table A.1). Since each prior is itself an ensemble, these 10 reconstructions effectively create an ensemble of ensembles. To minimize ambiguity, we will henceforth refer to the set of 10 reconstructions as the “multi-model ensemble”, and the DA ensemble for each individual reconstruction as a “prior/posterior ensemble”.

Forward model estimates of the NTREND records in each reconstruction are determined by applying equation A.2 to CRU-TS 4.01. We assess the skill of each reconstruction using time-series of mean Northern Hemisphere extratropical (30°N–90°N) MJJA temperature, instrumental spatial field grid points, and independent proxy records. The skill of the extratropical time series is determined using a Monte Carlo calibration-validation procedure (section A.8.1). Spatial skill is computed against the Berkeley Earth surface temperature field (BEST; Rohde et al., 2013) over the period 1901 - 1988 CE. The BEST instrumental record is not used in the forward model and localization calibrations, which instead leverage the CRU product. However, we caution that BEST is not a truly independent dataset, as both BEST and CRU are partly based on the same instrumental climate data. As an additional validation we assess the ability of DA to reconstruct withheld proxy time series. We perform a series of leave-one-out assimilations for each model by iteratively removing a single proxy time-series from the NTREND network and assimilating the remaining 53 records. In these experiments, we construct the prior from the average temperatures over the removed site’s optimal growing season at the grid point

closest to the removed site. This allows us to apply Equation A.2 to the posterior to estimate the removed record from the reconstruction. We then compare this estimate to the real withheld NTREND record.

We next calculate a mean reconstruction for the multi-model ensemble. To do so, we first calculate ensemble-mean values from the posterior of each of the reconstructions. The mean of the multi-model ensemble is then calculated as the mean of these 10 posterior ensemble means. We quantify uncertainty of the multi-model mean using first the mean of the 10 posterior ensemble widths:

$$\sigma_{\text{multi-model mean}}^2 = \frac{1}{10} \sum_{i=1}^{10} \sigma_{\text{posterior ensemble } i}^2 \quad (\text{A.5})$$

and then the  $2\sigma$  width of the multi-model ensemble for the series. We first determine the multi-model ensemble-mean for the extratropical MJJA time series. We next compute a mean spatial reconstruction for the multi-model ensemble by linearly interpolating each reconstruction to the lowest model resolution and averaging at each grid point.

We compare the multi-model mean spatial product to several recent temperature CFRs summarized in Table A.3. In brief, Guillet et al. (2017) focused on reconstructing high-frequency temperature anomalies associated with known volcanic eruptions using a network of a similar size and composition to the NTREND network in a linear regression framework and their work provides a comparison point with Anchukaitis et al. (2017). The LMR 2.1 reconstruction applied an offline EnSRF DA to the PAGES2k network and allows us to compare DA reconstructions using different proxy networks (Tardif et al., 2019). From Zhu et al. (2020), we examine the reconstruction of mean June through August (JJA) temperatures using PAGES2k trees. The Neukom et al. (2019) DA offers another comparison point, using a proxy network of intermediate size derived from a screened version of PAGES2k. Neukom et al. (2019) performed an ensemble of reconstructions using different methods and recommend using the ensemble mean reconstruction for climate analysis; however, we only focus on the DA product to emphasize the differences in reconstructions that arise when using similar methodologies.

We examine the temperature response to external forcing for both the reconstruction ensemble and temperature CFRs. We compare temperature anomalies between the Me-

dieval Climate Anomaly (MCA; 950 - 1250 CE) and the Little Ice Age (LIA; 1450 - 1850 CE) (Masson-Delmotte et al., 2013; Anchukaitis et al., 2017), and separately use superposed epoch analysis (Haurwitz and Brier, 1981) to determine composite mean responses to major tropical volcanic eruptions. For the volcanic events, we follow Sigl et al. (2015) and identify years containing a global eruption forcing magnitude equal to or larger than the 1884 Krakatoa eruption ( $n = 20$ ), which yields the following event years: 916, 1108, 1171, 1191, 1230, 1258, 1276, 1286, 1345, 1453, 1458, 1595, 1601, 1641, 1695, 1809, 1815, 1832, 1836, and 1884 CE (Sigl et al., 2015; Anchukaitis et al., 2017). We calculate temperature anomalies relative to the mean of the five years preceding each of these event years.

## **A.4 Results**

### **A.4.1 Pseudo-proxy Experiments**

The pseudo-proxy reconstructions are most skillful in the extratropical Northern Hemisphere (Figure A.2). In this region, ocean basin correlations are lower relative to land with notable exceptions over the eastern and north-western edges of the Pacific. Correlations generally decline with increasing distance from the extratropical Northern Hemisphere and the tree-ring network, although significant spatial heterogeneity exists throughout the tropics. The climate model covariance biases cause the largest reductions in correlation coefficients and sharply reduce skill outside of the extratropical Northern Hemisphere. Network attrition and proxy noise have comparatively minor effects over the full period. Results for other skill metrics show similar behavior (Figures A.11, A.12, and A.13).

We next compare the most realistic (biased-model, noisy-proxy, temporal-attrition) DA experiments to PPR reconstructions. Given the strict reconstruction radius in PPR, and the spatial pattern of DA skill, we consider only the extratropical Northern Hemisphere in our discussion. The skill metrics for the mean extratropical time series are similar for the two methods (Table A.7; Figures A.14, A.15). The regional spatial correlations of the DA and PPR reconstructions for the CESM and MPI targets (Figures A.3 and A.16, respectively) are also comparable: each exhibits correlations with the target

field greater than 0.7 in Scandinavia, western Siberia, and western Canada, and these regions correspond to the best coverage by the proxy network. Similarly, both methods exhibit low correlations in southeastern Canada, eastern Siberia, and in the region of the Black and Caspian Seas. The DA does however exhibit a broader spatial region of high correlation than PPR, and DA correlations are higher than PPR values at nearly all grid points. Similarly, DA reconstructions exhibit lower RMSE values at most grid points. Standard deviation ratios indicate that the DA reconstructions underestimate temporal temperature variability, but this effect is less severe near the proxy sites. In contrast with DA, PPR time series  $\sigma$  ratios neither strictly overestimate nor strictly underestimate temporal variability, instead demonstrating a mixed response over the hemisphere. In general, our DA reconstructions underestimate variability more strongly than the PPR analogues. Mean biases are comparable, with both methods exhibiting similar spatial patterns and bias magnitudes, although it is interesting to note that the spatial patterns of bias change markedly depending on the target field.

#### **A.4.2 Real NTREND Reconstruction**

For the real NTREND data assimilation, validation statistics for the mean extratropical MJJA time series are similar across all priors (Table A.2) with mean correlations of 0.70, RMSE of 0.19 °C, and absolute mean bias of 0.06 °C. Temporal variability is close to the target with mean standard deviation ratios of 1.11. Time series obtained using different model priors (Figure A.17) have a mean range of 0.22 °C over the period of full coverage (1750-2988 CE;  $n = 54$ ). However, the reconstructed time series diverge as the network becomes sparse, with a range of 0.76 °C by the first year of the reconstruction (750 CE;  $n = 4$ ). The model ensemble-mean time series exhibits similar skill values as the reconstructions for the individual models (Table A.2) with a correlation of 0.72, RMSE of 0.18 °C, temporal  $\sigma$  ratio of 1.06, and a mean bias of 0.05 °C.

We compare the extratropical MJJA time series for the multi-model mean to analogous time series extracted from the Berkeley Earth (BEST) instrumental record and the Anchukaitis et al. (2017) NTREND PPR reconstruction (Figure A.4). The DA series shows similar behavior to BEST from 1880-1988 CE, although both the DA and PPR

reconstructions of Anchukaitis et al. (2017) diverge from this dataset over the earliest period from 1850-1879 CE. This may reflect a warm bias (Parker, 1994; Frank et al., 2007; Böhm et al., 2010) and limited spatial coverage (Rohde et al., 2013; Anchukaitis et al., 2017) in the early instrumental temperature record. The DA and PPR time series show similar behavior over most of the record, with a correlation coefficient of 0.88. Temporal variability is generally higher in the PPR series than in the DA. Prior to about 1100 CE, the series' running standard deviations show larger differences, which is caused by the decrease in DA reconstructed variability.

Most spatial validation statistics show similar patterns to those observed in the pseudo-proxy experiments (Figure A.5). Correlation coefficients and standard deviation ratios indicate the highest skill over Scandinavia, central and northern Asia, and northwestern North America, the regions of densest network coverage. Correlation coefficients approach 0.8 and standard deviation ratios approach 1 near the proxy sites themselves. Over land, mean biases are typically below 0.5 °C, with the largest largest over central Canada and eastern Siberia and smallest over the Arctic Archipelago, Alaska, and west-central Asia. Away from the proxy sites, temporal variability is underestimated, particularly over the oceans. However, most land grid points exhibit  $\sigma$  ratios near 1 with a slight overestimate in central Asia and northern Japan. Much of the temporal variability in the extratropical mean time series is driven by land grid points, and this tendency helps reconcile Figure A.5 with extratropical mean time series  $\sigma$  ratios near 1. RMSE values are typically less than 0.6 °C, but rise to values near 1 °C over the North Pacific, central Canada, and north of the Caspian Sea.

Independent proxy validation statistics (Table A.4) show median correlation coefficients near 0.5, and RMSE values near 1°C. Temporal variability is underestimated relative to the target series with  $\sigma$  ratios typically between 0.3 and 0.4. Mean biases are variable and depend on the prior model used. Not surprisingly given the sparsity of the NTREND network, removing even a single proxy record from the assimilation can substantially reduce the ability to reconstruct temperature anomalies at nearby grid cells. Consequently, the leave-one-out assimilation process we use to assess independent proxy skill almost certainly underestimates overall field validation skill. Nevertheless, these

values are comparable to previous efforts with median correlation coefficients somewhat higher than those in Hakim et al. (2016) and Tardif et al. (2019).

### **A.4.3 Epochal Temperature Changes**

We next examine the temperature change between the Medieval Climate Anomaly (MCA; 950 - 1250 CE) and the Little Ice Age (LIA; 1450 - 1850 CE) (Masson-Delmotte et al., 2013; Anchukaitis et al., 2017). The reconstructions nearly all indicate warmer temperatures during the MCA throughout the high latitudes with maximum anomalies typically over northeastern Canada (Figure A.6). However, anomaly magnitudes vary across reconstructions with values ranging from over 1.6 °C (for CCSM4, MIROC, MPI priors) to less than 0.8 °C (IPSL and FGOALS priors). The spatial pattern also varies by model prior. Many reconstructions show stronger anomalies in Fennoscandia, northeastern Asia, and northwestern North America, but these patterns do not occur in all models.

Comparing the MCA-LIA difference for our multi-model mean reconstruction with other CFRs (Figure A.7), we find our spatial anomaly patterns most similar to Anchukaitis et al. (2017). Anomaly magnitudes are also comparable, except over northeastern Canada. In the Anchukaitis et al. (2017) reconstruction, this region exhibits anomalously high medieval temperatures ( $> 3$  °C), which they attribute to a detrending artifact in a tree-ring record from Quebec. By contrast, our DA reconstruction produces a maximum medieval anomaly of 1 °C for this region, in better agreement with other proxy reconstructions (e.g. 0-1.5°C; Sundqvist et al., 2014). Comparing the results of this study to Neukom et al. (2019), we observe that both NTREND DA and Neukom et al. (2019) exhibit a positive anomaly over most of the high-latitude Northern Hemisphere; however, the anomalies in the Neukom et al. (2019) product have much larger magnitudes and the maxima of the North America features occur in different locations. Zhu et al. (2020) also indicate positive anomalies in the Northern Hemisphere, but these are lower magnitude than the other products and more spatially localized. By contrast, the LMR2.1 product (Tardif et al., 2019) exhibits an anomaly pattern notably different from the other reconstructions, with a strong positive anomaly in the Arctic Ocean north of Siberia. Since the Guillet et al. (2017) reconstruction reflects high-pass filtered reconstructed temperatures, we do

not consider it in this comparison.

#### **A.4.4 Volcanic Response**

We next examine the composite mean response to major tropical volcanic eruptions. Our 10 reconstructions show broadly similar responses to large tropical volcanic eruptions (Figure A.8), with the spatial pattern characterized by a strong cold anomaly in northern Canada and a second region of cooling extending from Fennoscandia east of the Caspian Sea toward central Asia. However, the extent and magnitude of these vary between the different reconstructions. Several regions also exhibit markedly different spatial patterns across the 10 reconstructions. In particular, the response in central North America and eastern Asia appears highly sensitive to the choice of model prior.

Comparing the volcanic pattern for our multi-model mean reconstruction with the other existing CFRs (Figure A.9) shows large differences in spatial patterns, magnitudes, and even sign of the anomalies. In general, most CFRs show some combination of cooling anomalies in northern North America and northern Asia, with a slight neutral or warming anomaly in the North Pacific. However, these features are not present in all the CFRs and vary in maximum magnitude. The mean of our model ensemble, Anchukaitis et al. (2017), and Guillet et al. (2017) products all exhibit the northern Canada and western Asia cooling features and the spatial extent is similar for the two NTREND products. In contrast, the Guillet et al. (2017) Canadian feature is centered farther east, and its northern Asian feature is stronger (near 1.5 °C) with a maximum more strongly localized to northern Siberia. These two features are also present in Zhu et al. (2020), but maximum cooling is smaller in magnitude. The LMR2.1 does not show distinct north Asian terrestrial cooling, although an anomaly of 0.6 C is reconstructed in the Arctic Ocean north of Siberia. This reconstruction also demonstrates a North American response pattern similar to Zhu et al. (2020) with a reduced magnitude of cooling in northern Canada. The Neukom et al. (2019) product again shows the largest anomalies, with values greater than 1.5 °C over much of northern Siberia and Fennoscandia. This feature does not extend as far south as in the NTREND DA ensemble-mean but is zonally wider. Neukom et al. (2019) also show a single strong North American feature with cooling magnitudes near 1.2 °C. Inter-

estingly, Neukom et al. (2019) exhibits a North Pacific warming response that strengthens one year after the volcanic event, a feature also evident in the Anchukaitis et al. (2017) reconstruction that may reflect changes in atmospheric circulation following an eruption (e.g. Robock, 2000; Stenchikov et al., 2006; Christiansen, 2008; Schneider et al., 2009)

## A.5 Discussion

The pseudo-proxy experiments indicate that regions of high reconstruction skill for the assimilated NTREND network is limited to the extratropical Northern Hemisphere when using biased climate model priors. This finding supports work by Franke et al. (2020) and suggests that analyses of temperatures using the NTREND network should be limited to this region, consistent with Wilson et al. (2016) and Anchukaitis et al. (2017). In comparison with Anchukaitis et al. (2017) (NTREND PPR), our DA method exhibits similar skill at reconstructing mean Northern Hemisphere extratropical MJJA time series using the NTREND network, but also provides continuous field estimates of past temperature and improves the spatial correlation and RMSE. We suggest this improvement arises at least in part from the contrast between PPR’s strict-limited search radius and the DA’s longer localization radii. Many NTREND sites exhibit statistically significant covariance with the MJJA temperature field outside of PPR’s 2000 km maximum search radius (see Figure 5 of Anchukaitis et al. (2017)), and these distal covariances are not used to improve the PPR reconstruction. By contrast, the DA uses no localization in these pseudo-proxy experiments (Table A.6) and if the model prior provides a good estimate of a proxy site’s field covariance, the proxy record can inform the reconstruction of distal grid points. Ultimately, these results suggest that our DA method improves on the spatial component of Anchukaitis et al. (2017) for reconstructing a Northern Hemisphere temperature history of the Common Era from the NTREND network. We note that, as is the case for most field reconstruction methods (Ammann and Wahl, 2007; Tingley et al., 2012), our offline DA method implicitly assumes the broad-scale covariance patterns can be considered stationary through time. Transient offline (e.g. Bhend et al., 2012; Valler et al., 2019; Franke et al., 2020) or online assimilation techniques (e.g. Perkins and Hakim, 2017) may offer

additional improvements.

Our results also highlight the sensitivity of the DA reconstructions to the model prior. In the pseudo-proxy experiments, the introduction of model covariance bias reduces widespread global skill to the high latitude Northern Hemisphere and the regions nearest the proxy sites. Network attrition and proxy noise cause comparatively small effects over the full period, a finding in agreement with Dee et al. (2016). Given this potential for perfect-model experiments to exaggerate the magnitude and spatial extent of DA skill, we encourage future DA proof-of-concept and sensitivity studies to consider perfect-model experiments in conjunction with biased-model cases. In contrast with these results, previous assimilation efforts have found little sensitivity to the choice of prior (Hakim et al., 2016). The small size of the NTREND network may exacerbate this sensitivity, but even assimilations using larger networks may be sensitive to the choice of priors in those periods with reduced proxy coverage.

Reconstructions are most sensitive to the prior when the proxy network becomes small. For example, despite using the same proxy network and reconstruction technique, mean extratropical MJJA temperature time series diverge by more than 0.5 °C in the earliest parts of the reconstruction when the number of sites in our network is limited (Figure A.17). The use of different priors also produces noticeable differences in spatial MCA-LIA temperature anomaly patterns (Figure A.6), which we interpret as arising from the reduced size of the proxy network during the MCA. In contrast, the volcanic response maps present a more consistent spatial pattern (Figure A.8), which we attribute to the larger size of the proxy network during most of the volcanic events. The magnitude of the forced response may also contribute to similarity across the priors; however, the volcanic response maps still exhibit different spatial patterns in regions like east Asia where the proxy network is sparse.

The consistency with which the DA underestimates the temporal variability of the target field, particularly over the oceans and far from the proxy sites, requires consideration. In this study, we focus on time series derived from the posterior ensemble-mean at each time step. However, this focus on the ensemble-mean neglects the width of the full posterior ensemble. Like many offline EnSRF studies (e.g. Hakim et al., 2016; Dee

et al., 2016; Steiger et al., 2018), our method uses a stationary prior in each time step; thus, the prior ensemble-mean is constant through time. As the proxy network becomes sparse, update magnitudes decrease, and the posterior ensemble more closely resembles the prior. When this occurs, the reconstructed ensemble-mean time series will closely resemble the mean of the prior ensemble, and the time series' temporal variability will approach zero. Similarly, regions far from the proxy network will exhibit smaller update magnitudes, so grid point time series far from the proxy sites have lower  $\sigma$  ratios. However, this reduction in temporal variability is balanced by increased posterior ensemble width, which will remain near the spread of the prior ensemble. Incorporating the width of the posterior with ensemble-mean time series can produce a range that encompasses target time-series variability, but it is not always clear how to use these ranges in spatiotemporal analyses. Hence, we emphasize that users of DA products with constant priors should carefully consider how changes in the proxy network affect the temporal variability of posterior ensemble-mean time series and make use of the posterior range when possible. We also note that allowing the model prior to vary in each time step may help mitigate these effects, which again may argue for expanded future use of transient offline priors (e.g. Bhend et al., 2012; Valler et al., 2019; Franke et al., 2020) or online assimilation techniques (e.g. Perkins and Hakim, 2017) where possible.

The prior sensitivity and temporal variability effects underscore the importance of understanding how the proxy network affects the quality of the reconstruction (Esper et al., 2005; Wang et al., 2014). A key feature of DA techniques is the ability to estimate reconstruction uncertainty in each time step from the width of the posterior ensemble. Figure A.10 provides an example of such an analysis for the multi-model mean by examining the temperature response following the 1257 CE (Lavigne et al., 2013) and 1600 CE (De Silva and Zielinski, 1998) volcanic eruptions in conjunction with the full posterior width. The uncertainty maps for both events show maxima in central North American and northeastern Asia and suggest that associated temperature anomalies should be interpreted more cautiously. Notably, these regions correspond to areas that are also sensitive to the prior in Figure A.8. By contrast, central and east-central Asia, Fennoscandia, central Europe, and southwestern Canada exhibit a narrow posterior for both events, so

volcanic anomalies in these regions are better constrained. Interestingly, the temperature response in 1601 CE is relatively small over much of central Europe and reconstruction uncertainty is relatively low, which suggests this feature may be a robust feature of the post-eruption climate anomaly. In addition to supporting analysis of reconstructed climate features, these uncertainty estimates can help identify regions that would benefit from increased network density (Comboul et al., 2015). In particular, we observe that northern North America and eastern Siberia would benefit from the development of new millennial-length temperature-sensitive tree-ring records.

The CFR comparison reveals the highly variable nature of spatial patterns and magnitudes of reconstructed temperature anomalies that result from different selections of proxy networks, target fields, and reconstruction methodologies. For example, despite using the same proxy network and target field, the DA multi-model mean and PPR result from Anchukaitis et al. (2017) have MCA-LIA anomalies that differ by over 2 °C in northeastern Canada (Figure A.7), which relates to the outsized effect of the Quebec tree-ring width record (Gennaretti et al., 2014) on the Anchukaitis et al. (2017) reconstruction. We note that the localization radii used in our reconstructions ( $\geq 9500$  km) allow proxies to influence grid cells farther away than the maximum 2000 km search radius used by Anchukaitis et al. (2017), so distant proxies are able to counter the effects of the Quebec record in the DA. Even within the same DA framework, our results indicate that reconstructed temperature responses are highly variable, particularly for MCA-LIA anomalies. These differences result from targeting different fields and leveraging different proxy networks. Aside from spatial and temporal coverage, we note that using proxy records that are not strictly temperature sensitive can introduce structural biases relative to other temperature CFRs. For example, the LMR2.1 reconstruction includes proxies that are sensitive to more than just temperature, which could possibly reduce update magnitudes and help explain the smaller magnitudes of the volcanic responses. Similarly, the Neukom et al. (2019) DA product and LMR2.1 incorporate proxies like corals and lake-sediments that are not present in the tree-ring based CFRs, and it is possible that these records influence the large magnitudes of the Neukom et al. (2019) DA climate responses or the atypical LMR2.1 MCA-LIA spatial pattern. However, we emphasize that these hy-

potheses are strictly speculative at this moment and that the differences in reconstructed climate response by themselves do not indicate whether one proxy network or reconstruction is superior to another in representing past climate variability. Instead, our CFR comparison highlights that, despite the recent decades of progress in understanding both methods and paleoclimate data (Hughes and Ammann, 2009; Frank et al., 2010; Smerdon et al., 2011; Tingley et al., 2012; Wang et al., 2014; Smerdon and Pollack, 2016; Christiansen and Ljungqvist, 2017; Esper et al., 2018), differences in reconstructions of past temperature still arise when using different proxy networks, different target seasons, and making different reconstruction choices, and these differences fundamentally influence our interpretation of the temperature response to radiative forcing (c.f. Wang et al., 2015). This observation calls for a revival of paleo-reconstruction intercomparison projects (e.g. Ammann, 2008; Graham and Wahl, 2011; Anchukaitis and McKay, 2015) in order to examine the behavior, strengths, and weaknesses of different proxy networks and reconstruction choices in a systematic and community-driven manner. Furthermore, such an effort would help identify regions with consistently large reconstruction uncertainties and indicate where to prioritize the development of new or the extension of existing tree-ring records.

## **A.6 Conclusions**

In this study, we assimilate a small but highly temperature-sensitive tree-ring network based on expert assessment to reconstruct summer (MJJA) temperature anomalies from 750-1988 CE. Our method is skillful in the extratropical Northern Hemisphere and improves on a previous spatial reconstruction using the same network, thereby providing a new dataset with which to examine temperature dynamics and climate response to radiative forcing over the last millennium. In a set of pseudo-proxy experiments, we find that our method is sensitive to climate model biases, so we perform an ensemble of reconstructions using 10 different climate model priors. Reconstructed temperature anomalies are sensitive to the selection of the model prior when the proxy network becomes sparse, but the reconstructed spatial patterns and time series converge to consistent values as the

number of sites in the NTREND proxy network increases. As one consequence of using static offline priors, our method underestimates temporal variability particularly when the proxy network becomes small, which argues for the future use of transient offline priors, online assimilation techniques in DA paleoclimate reconstructions, and expanded proxy development. There is also a need for continued development of proxy system forward models, particularly for the important MXD metric. The influence of the proxy network coverage on the reconstructions emphasizes the importance of analyzing reconstructed temperature anomalies in conjunction with estimates of their uncertainty. These uncertainty estimates emerge naturally for both spatial fields and time series from the DA posterior ensembles and are an enhancement over previous reconstructions using the NTREND dataset. In addition to gauging reconstruction validity, the uncertainty estimates identify regions that would benefit from additional proxy records and support the development of more millennial-length temperature-sensitive tree-ring records in treeline North America and eastern Siberia especially. Comparison of our reconstruction with other temperature CFRs indicates that reconstructed temperature anomalies have highly variable spatial patterns and magnitudes, even within similar reconstruction frameworks and proxy network. These different climate responses call for a renewed paleo-reconstruction intercomparison framework in which to systematically examine the effects of network selection across reconstruction techniques and prioritize regions for future record development.

### **Data availability**

The NTREND proxy data and the earlier reconstructions are available from the NOAA NCEI World Data Service for Paleoclimatology (<https://www.ncdc.noaa.gov/paleo-search/study/19743>). The NTREND-DA ensemble reconstructions will be available from NOAA NCEI World Data Service for Paleoclimatology (<https://www.ncei.noaa.gov/access/paleo-search/study/33632>). Model priors from the CMIP5 and CESM LME are available on the Earth System Grid (<https://esgf-node.llnl.gov/projects/esgf-llnl/>) and the NCAR Climate Data Gateway (<https://www.earthsystemgrid.org/>), respectively. The data and

code used to run these analyses and a function reproducing the results and figures from this paper are available at <https://doi.org/10.5281/zenodo.3989941>.

### **Acknowledgements**

The authors acknowledge support from the Climate Program Office of the National Oceanographic and Atmospheric Administration (NOAA grants NA18OAR4310420 to KJA, NA18OAR4310426 to JEG and FZ, and NA18OAR4310422 to GJH). GJH also acknowledges support from the NSF through grant AGS-1702423. JMK and KJA were supported by NSF grant AGS-1803946. JET and JMK acknowledge support from NSF grant #AGS-1602301 and Heising-Simons Foundation grant #2016-05. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table A.1 of this paper) for producing and making available their model output.

## A.7 References

- Acevedo, W., Fallah, B., Reich, S., and Cubasch, U.: Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model, *Climate of the Past*, 13, 545–557, 2017.
- Ammann, C.: The Paleoclimate reconstruction challenge, *PAGES News*, 16, 4, 2008.
- Ammann, C. M. and Wahl, E. R.: The importance of the geophysical context in statistical evaluations of climate reconstruction procedures, *Climatic Change*, 85, 71–88, 2007.
- Anchukaitis, K. J. and McKay, N.: PAGES2k: Advances in climate field reconstructions, *PAGES Magazine*, 22, 98, 2015.
- Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D’Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B. E., et al.: Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions, *Quaternary Science Reviews*, 163, 1–22, 2017.
- Andrews, A.: A square root formulation of the Kalman covariance equations., *AIAA Journal*, 6, 1165–1166, 1968.
- Annan, J. and Hargreaves, J.: Identification of climatic state with limited proxy data, *Climate of the Past*, 8, 1141–1151, 2012.
- Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, *Climate of the Past*, 8, 963–976, 2012.
- Björklund, J., Gunnarson, B. E., Seftigen, K., Esper, J., et al.: Blue intensity and density from northern Fennoscandian tree rings, exploring the potential to improve summer temperature reconstructions with earlywood information, *Climate of the Past*, 10, 877–885, 2014.
- Böhm, R., Jones, P. D., Hiebl, J., Frank, D., Brunetti, M., and Maugeri, M.: The early instrumental warm-bias: a solution for long central European temperature series 1760–2007, *Climatic Change*, 101, 41–67, 2010.

- Chen, Z. et al.: Bayesian filtering: From Kalman filters to particle filters, and beyond, *Statistics*, 182, 1–69, 2003.
- Christiansen, B.: Volcanic eruptions, large-scale modes in the Northern Hemisphere, and the El Niño–Southern Oscillation, *Journal of Climate*, 21, 910–922, 2008.
- Christiansen, B. and Ljungqvist, F. C.: Challenges and perspectives for large-scale temperature reconstructions of the past two millennia, *Reviews of Geophysics*, 55, 40–96, 2017.
- Christiansen, B., Schmith, T., and Thejll, P.: A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness, *Journal of Climate*, 22, 951–976, 2009.
- Comboul, M., Emile-Geay, J., Hakim, G. J., and Evans, M. N.: Paleoclimate Sampling as a Sensor Placement Problem, *Journal of Climate*, 28, 7717–7740, <https://doi.org/10.1175/JCLI-D-14-00802.1>, 2015.
- Cook, E. R., Briffa, K. R., and Jones, P. D.: Spatial regression methods in dendroclimatology: a review and comparison of two techniques, *International Journal of Climatology*, 14, 379–402, 1994.
- Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K.: Drought reconstructions for the continental United States, *Journal of Climate*, 12, 1145–1162, 1999.
- Cook, E. R., Anchukaitis, K. J., Buckley, B. M., D’Arrigo, R. D., Jacoby, G. C., and Wright, W. E.: Asian monsoon failure and megadrought during the last millennium, *Science*, 328, 486–489, 2010a.
- Cook, E. R., Seager, R., Heim Jr, R. R., Vose, R. S., Herweijer, C., and Woodhouse, C.: Megadroughts in North America: Placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context, *Journal of Quaternary Science*, 25, 48–61, 2010b.

- Cook, E. R., Krusic, P. J., Anchukaitis, K. J., Buckley, B. M., Nakatsuka, T., Sano, M., et al.: Tree-ring reconstructed summer temperature anomalies for temperate East Asia since 800 CE, *Climate Dynamics*, 41, 2957–2972, 2013.
- Cort, G. D., Chevalier, M., Burrough, S. L., Chen, C. Y., and Harrison, S. P.: An uncertainty-focused database approach to extract spatiotemporal trends from qualitative and discontinuous lake-status histories, *Quaternary Science Reviews*, 258, 106 870, <https://doi.org/10.1016/j.quascirev.2021.106870>, 2021.
- Cowtan, K. and Way, R. G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, *Quarterly Journal of the Royal Meteorological Society*, 140, 1935–1944, 2014.
- De Silva, S. L. and Zielinski, G. A.: Global influence of the AD 1600 eruption of Huaynaputina, Peru, *Nature*, 393, 455–458, 1998.
- Dee, S. G., Steiger, N. J., Emile-Geay, J., and Hakim, G. J.: On the utility of proxy system models for estimating climate states over the Common Era, *Journal of Advances in Modeling Earth Systems*, 8, 1164–1179, 2016.
- Delworth, T. L. and Mann, M. E.: Observed and simulated multidecadal variability in the Northern Hemisphere, *Climate Dynamics*, 16, 661–676, 2000.
- Dubinkina, S. and Goosse, H.: An assessment of particle filtering methods and nudging for climate state reconstructions, *Climate of the Past*, 9, 1141–1152, 2013.
- Esper, J., Frank, D. C., Wilson, R. J., and Briffa, K. R.: Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophysical Research Letters*, 32, 2005.
- Esper, J., St George, S., Anchukaitis, K., D'Arrigo, R., Ljungqvist, F. C., Luterbacher, J., Schneider, L., Stoffel, M., Wilson, R., and Buntgen, U.: Large-scale, millennial-length temperature reconstructions from tree-rings, *Dendrochronologia*, 50, 81–90, <https://doi.org/10.1016/j.dendro.2018.06.001>, 2018.

- Evans, M., Kaplan, A., Cane, M., and Villalba, R.: Globality and optimality in climate field reconstructions from proxy data, in: *Interhemispheric Climate Linkages*, pp. 53–XV, Elsevier, 2001.
- Evans, M. N., Kaplan, A., and Cane, M. A.: Pacific sea surface temperature field reconstruction from coral  $\delta^{18}\text{O}$  data using reduced space objective analysis, *Paleoceanography*, 17, <https://doi.org/10.1029/2000PA000590>, 2002.
- Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Science Reviews*, 76, 16–28, 2013.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, 99, 10 143–10 162, 1994.
- Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343–367, 2003.
- Frank, D., Büntgen, U., Böhm, R., Maugeri, M., and Esper, J.: Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target, *Quaternary Science Reviews*, 26, 3298–3310, 2007.
- Frank, D., Esper, J., Zorita, E., and Wilson, R.: A noodle, hockey stick, and spaghetti plate: a perspective on high-resolution paleoclimatology, *Wiley Interdisciplinary Reviews: Climate Change*, 1, 507–516, 2010.
- Franke, J., Valler, V., Brönnimann, S., Neukom, R., and Jaume-Santero, F.: The importance of input data quality and quantity in climate field reconstructions – results from the assimilation of various tree-ring collections, *Climate of the Past*, 16, 1061–1074, <https://doi.org/10.5194/cp-16-1061-2020>, 2020.
- Fritts, H. C.: *Reconstructing large-scale climatic patterns from tree-ring data: a diagnostic analysis.*, University of Arizona Press, 1991.

- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757, 1999.
- Gennaretti, F., Arseneault, D., Nicault, A., Perreault, L., and Bégin, Y.: Volcano-induced regime shifts in millennial tree-ring chronologies from northeastern North America, *Proceedings of the National Academy of Sciences*, 111, 10 077–10 082, 2014.
- Gill, E. C., Rajagopalan, B., Molnar, P., and Marchitto, T. M.: Reduced-dimension reconstruction of the equatorial Pacific SST and zonal wind fields over the past 10,000 years using Mg/Ca and alkenone records, *Paleoceanography*, 31, 928–952, 2016.
- Goosse, H.: An additional step toward comprehensive paleoclimate reanalyses, *Journal of Advances in Modeling Earth Systems*, pp. n/a–n/a, <https://doi.org/10.1002/2016MS000739>, 2016.
- Goosse, H.: Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries, *Climate Dynamics*, 48, 1483–1501, 2017.
- Goosse, H., Guiot, J., Mann, M. E., Dubinkina, S., and Sallaz-Damaz, Y.: The Medieval Climate Anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation, *Global and Planetary Change*, 84, 35–47, 2012.
- Graham, N. and Wahl, E.: Paleoclimate reconstruction challenge, *PAGES/CLIVAR Newsletter*, 19, 71–72, 2011.
- Guillet, S., Corona, C., Stoffel, M., Khodri, M., Lavigne, F., Ortega, P., Eckert, N., Sieleenou, P. D., Daux, V., Churakova, O. v., et al.: Climate response to the Samalas volcanic eruption in 1257 revealed by proxy records, *Nature Geoscience*, 10, 123–128, 2017.
- Guillot, D., Rajaratnam, B., and Emile-Geay, J.: Statistical paleoclimate reconstructions via Markov random fields, *The Annals of Applied Statistics*, 9, 324–352, 2015.
- Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The Last Millennium Climate Reanalysis project: Framework

- and first results, *Journal of Geophysical Research: Atmospheres*, 121, 6745–6764, 2016.
- Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset, *International Journal of Climatology*, 34, 623–642, 2014.
- Haurwitz, M. W. and Brier, G. W.: A critique of the superposed epoch analysis method: its application to solar–weather relations, *Monthly Weather Review*, 109, 2074–2079, 1981.
- Hegerl, G. and Stott, P.: From past to future warming, *Science*, 343, 844–845, 2014.
- Hegerl, G. C., Hasselmann, K., Cubasch, U., Mitchell, J. F., Roeckner, E., Voss, R., and Waszkewitz, J.: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change, *Climate Dynamics*, 13, 613–634, 1997.
- Hegerl, G. C., Crowley, T. J., Baum, S. K., Kim, K.-Y., and Hyde, W. T.: Detection of volcanic, solar and greenhouse gas signals in paleo-reconstructions of Northern Hemispheric temperature, *Geophysical Research Letters*, 30, <https://doi.org/10.1029/2002GL016635>, 2003.
- Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., and Zorita, E.: Detection of human influence on a new, validated 1500-year temperature reconstruction, *Journal of Climate*, 20, 650–666, 2007.
- Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, 129, 123–137, 2001.
- Hughes, M. and Ammann, C.: The future of the past—an earth system framework for high resolution paleoclimatology: editorial essay, *Climatic Change*, 94, 247–259, 2009.

- Kaufman, D.: A Community-Driven Framework for Climate Reconstructions, *Eos, Transactions American Geophysical Union*, 95, 361–362, <https://doi.org/10.1002/2014eo400001>, 2014.
- Lavigne, F., Degeai, J.-P., Komorowski, J.-C., Guillet, S., Robert, V., Lahitte, P., Oppenheimer, C., Stoffel, M., Vidal, C. M., Pratomo, I., et al.: Source of the great AD 1257 mystery eruption unveiled, Samalas volcano, Rinjani Volcanic Complex, Indonesia, *Proceedings of the National Academy of Sciences*, 110, 16 742–16 747, 2013.
- Lean, J. L. and Rind, D. H.: How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006, *Geophysical Research Letters*, 35, <https://doi.org/10.1029/2008GL034864>, 2008.
- Liu, H., Liu, Z., and Lu, F.: A Systematic Comparison of Particle Filter and EnKF in Assimilating Time-Averaged Observations, *Journal of Geophysical Research: Atmospheres*, 122, 13–155, 2017.
- Mann, M. E. and Rutherford, S.: Climate reconstruction using ‘Pseudoproxies’, *Geophysical Research Letters*, 29, 139–1, 2002.
- Mann, M. E., Bradley, R. S., and Hughes, M. K.: Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779–787, 1998.
- Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F.: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, 326, 1256–1260, 2009.
- Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, *Ocean Modelling*, 5, 91–127, 2003.
- Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Rouco, J. G., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A.: Information from

- Paleoclimate Archives, in: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., pp. 383–464, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- Matsikaris, A., Widmann, M., and Jungclaus, J. H.: On-line and off-line data assimilation in palaeoclimatology: a case study, *Climate of the Past*, 11, 81–93, 2015.
- McCarroll, D., Pettigrew, E., Luckman, A., Guibal, F., and Edouard, J.-L.: Blue reflectance provides a surrogate for latewood density of high-latitude pine tree rings, *Arctic, Antarctic, and Alpine Research*, 34, 450–453, 2002.
- Meehl, G. A., Washington, W. M., Ammann, C. M., Arblaster, J. M., Wigley, T., and Tebaldi, C.: Combinations of natural and anthropogenic forcings in twentieth-century climate, *Journal of Climate*, 17, 3721–3727, 2004.
- Meko, D.: Dendroclimatic Reconstruction with Time Varying Predictor Subsets of Tree Indices, *Journal of Climate*, 10, 687–696, 1997.
- Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P.: No evidence for globally coherent warm and cold periods over the preindustrial Common Era, *Nature*, 571, 550–554, 2019.
- Okazaki, A. and Yoshimura, K.: Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction., *Climate of the Past*, 13, 379–393, 2017.
- Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., and Kosro, P. M.: Assimilation of surface velocity data into a primitive equation coastal ocean model, *Journal of Geophysical Research: Oceans*, 107, <https://doi.org/10.1029/2000JC000511>, 2002.
- Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G.: Climate variability and change since 850

- CE: An ensemble approach with the Community Earth System Model, *Bulletin of the American Meteorological Society*, 97, 735–754, 2016.
- PAGES2k Consortium: Continental-scale temperature variability during the past two millennia, *Nature Geoscience*, 6, 339–346, 2013.
- PAGES2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, *Scientific Data*, 4, 170 088, 2017.
- Parker, D.: Effects of changing exposure of thermometers at land stations, *International Journal of Climatology*, 14, 1–31, 1994.
- Perkins, W. A. and Hakim, G. J.: Reconstructing paleoclimate fields using online data assimilation with a linear inverse model, *Climate of the Past*, 13, 421–436, 2017.
- Phipps, S. J., McGregor, H. V., Gergis, J., Gallant, A. J., Neukom, R., Stevenson, S., Ackerley, D., Brown, J. R., Fischer, M. J., and Van Ommen, T. D.: Paleoclimate data–model comparison and the role of climate forcings over the past 1500 years, *Journal of Climate*, 26, 6915–6936, 2013.
- Robock, A.: Volcanic eruptions and climate, *Reviews of geophysics*, 38, 191–219, 2000.
- Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickhams, C., and Mosher, S.: Berkeley Earth Temperature Averaging Process, *Geoinformatics and Geostatistics: An Overview*, 1, 20–100, 2013.
- Rutherford, S., Mann, M., Delworth, T., and Stouffer, R.: Climate field reconstruction under stationary and nonstationary forcing, *Journal of Climate*, 16, 462–479, 2003.
- Rydval, M., Larsson, L.-Å., McGlynn, L., Gunnarson, B. E., Loader, N. J., Young, G. H., and Wilson, R.: Blue intensity for dendroclimatology: should we have the blues? Experiments from Scotland, *Dendrochronologia*, 32, 191–204, 2014.
- Schneider, D. P., Ammann, C. M., Otto-Bliesner, B. L., and Kaufman, D. S.: Climate response to large, high-latitude and low-latitude volcanic eruptions in the Community Climate System Model, *Journal of Geophysical Research: Atmospheres*, 114, 2009.

- Schneider, T.: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, *Journal of Climate*, 14, 853–871, 2001.
- Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F. B., and Phipps, S. J.: Separating Forced from Chaotic Climate Variability over the Past Millennium, *Journal of Climate*, 26, 6954–6973, <https://doi.org/10.1175/jcli-d-12-00826.1>, 2013.
- Schweingruber, F., Fritts, H., Bräker, O., Drew, L., and Schär, E.: The X-ray technique as applied to dendroclimatology, *Tree-Ring Bulletin*, 38, 61–91, 1978.
- Seager, R., Graham, N., Herweijer, C., Gordon, A. L., Kushnir, Y., and Cook, E.: Blueprints for Medieval hydroclimate, *Quaternary Science Reviews*, 26, 2322–2336, 2007.
- Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., et al.: Timing and climate forcing of volcanic eruptions for the past 2,500 years, *Nature*, 523, 543–549, 2015.
- Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudo-proxy experiments, *Wiley Interdisciplinary Reviews: Climate Change*, 3, 63–77, 2012.
- Smerdon, J. E. and Pollack, H. N.: Reconstructing Earth’s surface temperature over the past 2000 years: the science behind the headlines, *Wiley Interdisciplinary Reviews: Climate Change*, 7, 746–771, 2016.
- Smerdon, J. E., Kaplan, A., Zorita, E., González-Rouco, J. F., and Evans, M.: Spatial performance of four climate field reconstruction methods targeting the Common Era, *Geophysical Research Letters*, 38, <https://doi.org/10.1029/2011GL047372>, 2011.
- Solomon, A., Goddard, L., Kumar, A., Carton, J., Deser, C., Fukumori, I., Greene, A. M., Hegerl, G., Kirtman, B., Kushnir, Y., et al.: Distinguishing the roles of natural and anthropogenically forced decadal climate variability: implications for prediction, *Bulletin of the American Meteorological Society*, 92, 141–156, 2011.

- Steiger, N. J. and Smerdon, J. E.: A pseudoproxy assessment of data assimilation for reconstructing the atmosphere–ocean dynamics of hydroclimate extremes., *Climate of the Past*, 13, 1435–1449, 2017.
- Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of time-averaged pseudoproxies for climate reconstruction, *Journal of Climate*, 27, 426–441, 2014.
- Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, *Scientific Data*, 5, <https://doi.org/10.1086/sdata.2018.86>, 2018.
- Stenchikov, G., Hamilton, K., Stouffer, R. J., Robock, A., Ramaswamy, V., Santer, B., and Graf, H.-F.: Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models, *Journal of Geophysical Research: Atmospheres*, 111, 2006.
- Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Cruieger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., et al.: Atmospheric component of the MPI-M Earth system model: ECHAM6, *Journal of Advances in Modeling Earth Systems*, 5, 146–172, 2013.
- Stott, P. A. and Jones, G. S.: Variability of high latitude amplification of anthropogenic warming, *Geophysical Research Letters*, 36, <https://doi.org/10.1029/2009GL037698>, 2009.
- Stott, P. A. and Tett, S. F.: Scale-dependent detection of climate change, *Journal of Climate*, 11, 3282–3294, 1998.
- Stott, P. A., Gillett, N. P., Hegerl, G. C., Karoly, D. J., Stone, D. A., Zhang, X., and Zwiers, F.: Detection and attribution of climate change: a regional perspective, *Wiley Interdisciplinary Reviews: Climate Change*, 1, 192–211, 2010.
- Sundqvist, H. S., Kaufmann, D., MaKay, N., Axford, Y., Balascio, N., Briner, J. P., Cwynar, L., Sejrup, H. P., Subetto, D., Andrews, J., et al.: Arctic Holocene proxy

- climate database—new approaches to assessing geochronological accuracy and encoding climate variables, *Climate of the Past*, 10, 1605–1631, <https://doi.org/10.5194/cp-10-1605-2014>, 2014.
- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling., *Climate of the Past*, 15, 1251–1273, 2019.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93, 485–498, 2012.
- Tingley, M. P., Craigmiles, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: Statistical insights into paleoclimatic reconstructions, *Quaternary Science Reviews*, 35, 1–22, 2012.
- Valler, V., Franke, J., and Brönnimann, S.: Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation, *Climate of the Past*, 15, 1427–1441, <https://doi.org/10.5194/cp-15-1427-2019>, 2019.
- Van der Schrier, G. and Barkmeijer, J.: Bjerknes’ hypothesis on the coldness during AD 1790–1820 revisited, *Climate Dynamics*, 25, 537–553, 2005.
- Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J. E., and Rajaratnam, B.: Evaluating climate field reconstruction techniques using improved emulations of real-world conditions, *Climate of the Past*, 10, 1–19, <https://doi.org/10.5194/cp-10-1-2014>, 2014.
- Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., and Rajaratnam, B.: Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation, *Geophysical Research Letters*, 42, 7162–7170, 2015.
- Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, 130, 1913–1924, 2002.

- Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., and Barkmeijer, J.: Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium, *Climate of the Past*, 6, 627–644, <https://doi.org/10.5194/cp-6-627-2010>, 2010.
- Wikle, C. K. and Berliner, L. M.: A Bayesian tutorial for data assimilation, *Physica D: Nonlinear Phenomena*, 230, 1–16, 2007.
- Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D’arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., et al.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, *Quaternary Science Reviews*, 134, 1–18, 2016.
- Wilson, R., Anchukaitis, K., Andreu-Hayles, L., Cook, E., D’Arrigo, R., Davi, N., Haberbauer, L., Krusic, P., Luckman, B., Morimoto, D., et al.: Improved dendroclimatic calibration using blue intensity in the southern Yukon, *The Holocene*, 29, 1817–1830, 2019.
- Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., and Anchukaitis, K. J.: Resolving the differences in the simulated and reconstructed temperature response to volcanism, *Geophysical Research Letters*, 47, e2019GL086908, 2020.
- Zorita, E., González-Rouco, F., and Legutke, S.: Testing the approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model, *Journal of Climate*, 16, 1378–1390, 2003.

## **A.8 Supplemental Information**

### **A.8.1 Data Assimilation Methods**

#### **The Ensemble Kalman Filter**

Our data assimilation method uses an ensemble Kalman filter approach (Evensen, 1994; Steiger et al., 2014; Hakim et al., 2016) to solve the update equation:

$$\mathbf{X}_a = \mathbf{X}_p + \mathbf{K}(\mathbf{Y} - \mathbf{Y}_e) \quad (\text{A.6})$$

in each reconstructed annual time step. Here  $\mathbf{X}_p$  is an initial ensemble of plausible climate states, an  $n \times m$  matrix where  $n$  is the number of state variables and  $m$  is the number of ensemble members.  $\mathbf{X}_a$  is the updated ensemble (the analysis), also an  $n \times m$  matrix.  $\mathbf{Y}$  is a  $d \times m$  matrix of observed proxy values, where  $d$  is the number of available proxy records in a given time step.  $\mathbf{Y}_e$  is a  $d \times m$  matrix consisting of model estimates of the proxy values. Each row  $y_{e_j}$  is determined by applying the forward model for the  $j^{\text{th}}$  proxy site to the ensemble via Equation A.2.  $\mathbf{K}$  is the Kalman Gain, an  $n$  by  $d$  matrix that weights the covariance of proxy sites with the target field by the uncertainties in the proxy observations and estimates.

We use an EnKF variant known as the ensemble square root Kalman filter (EnSRF; Andrews, 1968), which removes the need for perturbed observations (Whitaker and Hamill, 2002). Consequently,  $\mathbf{Y}$  is a matrix with constant rows. In the EnSRF formulation, ensemble deviations are updated separately from the mean, as per:

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_p + \mathbf{K}(\bar{\mathbf{y}} - \bar{\mathbf{y}}_e) \quad (\text{A.7})$$

$$\mathbf{X}'_a = \mathbf{X}'_p - \tilde{\mathbf{K}}\mathbf{Y}'_e \quad (\text{A.8})$$

where an overbar ( $\bar{\mathbf{x}}$ ) denotes an ensemble average, and a tick ( $\mathbf{X}'$ ) indicates deviations from an ensemble mean. Here, the ensemble mean is updated via the Kalman gain ( $\mathbf{K}$ ):

$$\mathbf{K} = \text{cov}(\mathbf{X}_p, \mathbf{Y}_e) \times [\text{cov}(\mathbf{Y}_e, \mathbf{Y}_e) + \mathbf{R}]^{-1} \quad (\text{A.9})$$

and the deviations are updated via an adjusted gain ( $\tilde{\mathbf{K}}$ ):

$$\tilde{\mathbf{K}} = \text{cov}(\mathbf{X}_p, \mathbf{Y}_e) \times [(\sqrt{\text{cov}(\mathbf{Y}_e, \mathbf{Y}_e) + \mathbf{R}})^{-1}]^T [\sqrt{\text{cov}(\mathbf{Y}_e, \mathbf{Y}_e) + \mathbf{R}} + \sqrt{\mathbf{R}}]^{-1} \quad (\text{A.10})$$

Here,  $\mathbf{R}$  denotes the observation error-covariance matrix ( $d \times d$ ). We do not consider correlated measurement errors in this study, so  $\mathbf{R}$  is a diagonal matrix whose elements

are the observation uncertainties determined from the variances of the residuals for the forward model regressions.

### Covariance Localization

We implement a covariance localization scheme, modifying the Kalman Gain equations to:

$$\mathbf{K} = \mathbf{W}_{\text{loc}} \circ \text{cov}(\mathbf{X}_{\text{p}}, \mathbf{Y}_{\text{e}}) \times [\mathbf{Y}_{\text{loc}} \circ \text{cov}(\mathbf{Y}_{\text{e}}, \mathbf{Y}_{\text{e}}) + \mathbf{R}]^{-1} \quad (\text{A.11})$$

and

$$\tilde{\mathbf{K}} = \mathbf{W}_{\text{loc}} \circ \text{cov}(\mathbf{X}_{\text{p}}, \mathbf{Y}_{\text{e}}) \times [(\sqrt{\mathbf{Y}_{\text{loc}} \circ \text{cov}(\mathbf{Y}_{\text{e}}, \mathbf{Y}_{\text{e}}) + \mathbf{R}})^{-1}]^{\text{T}} [\sqrt{\mathbf{Y}_{\text{loc}} \circ \text{cov}(\mathbf{Y}_{\text{e}}, \mathbf{Y}_{\text{e}}) + \mathbf{R}} + \sqrt{\mathbf{R}}]^{-1}. \quad (\text{A.12})$$

Here,  $\mathbf{W}_{\text{loc}}$  ( $n \times d$ ) and  $\mathbf{Y}_{\text{loc}}$  ( $d \times d$ ) are matrices of covariance localization weights applied to the covariance of proxy sites with model grid cells ( $\mathbf{W}_{\text{loc}}$ ) and proxy sites with one another ( $\mathbf{Y}_{\text{loc}}$ ). We implement localization weights as a fifth order Gaspari-Cohn polynomial (Gaspari and Cohn, 1999) applied to the distance between proxy sites and model grid cells ( $\mathbf{W}_{\text{loc}}$ ) or proxy sites with one another ( $\mathbf{Y}_{\text{loc}}$ ). Weights are applied to covariance matrices via element-wise multiplication.

The localization radius is an important free parameter that must be assessed independently for different model priors, reconstruction targets, and proxy networks. Here, we select localization radii using a two step process. For a given model prior and target field, we first assimilate the proxy network from 1901-1988 CE using each localization radius from 250 km to 50,000 km in steps of 250 km and a run with no localization. We then determine the  $\sigma$  ratio of each reconstructed extratropical MJJA time series in a calibration interval. We find the  $\sigma$  ratio closest to 1 and record the associated localization radius as “optimal”. We then calculate skill metrics for the extratropical MJJA time series over a validation interval using the reconstruction with the optimal radius.

To limit the sensitivity of this method to the calibration period (Christiansen et al., 2009), we perform this optimization using each set of 44 contiguous years from 1901-1988 CE once as a calibration interval and once as a validation interval. The final localization radius is the median of the 88 “optimal” radii, and the median validation skill metrics are reported.

### **Selection Criterion**

In the development of this method, we tested an RMSE selection criterion in addition to  $\sigma$  ratios. We find that correlation coefficients, RMSE values, and mean biases of the reconstructed mean extratropical MJJA time series are all insensitive to the choice of selection criteria (Table A.2, Table A.5), but that  $\sigma$  ratios are more sensitive. Specifically, mean  $\sigma$  ratios are near 0.8 for the RMSE selection criterion, but rise to 1.11 for the  $\sigma$  ratio scheme. Since the  $\sigma$  ratio localization selection criteria brings the  $\sigma$  ratio skill metric closer to 1 without appreciably altering the other skill metrics, and because of the tendency for our DA method to underestimate temporal variability, we use a  $\sigma$  ratio selection criterion.

Table A.1: Summary of climate models used to construct data assimilation prior ensembles. Climate models are listed along with the identifying acronym used in this study. The years of available output are provided with the experiment used to generate them. The size of the model prior generated from these years is also provided. Taylor et al. (2012) provide more details on the PMIP3 and CMIP5 experiments, and Otto-Bliesner et al. (2016) describe the LME.

Model	Acronym	Years: Experiment	Sample size ( $m$ )
BCC-CSM1-1	BCC	850-2000: past1000	1151
CCSM4	CCSM4	850-1850: past1000 1851-2005: historical	1156
CESM1.1-CAM5	CESM	850-2005: LME full-forcing	1156
CSIRO-Mk3L-1-2	CSIRO	851-1850: past1000 1851-2000: historical	1150
FGOALS-g1	FGOALS	1000-1999: past1000	1000
HadCM3	HadCM3	850-1850: past1000 1859-2000: historical	1147
IPSL-CM5A-LR	IPSL	850-1850: past1000 1851-2005: historical	1156
MIROC-ESM	MIROC	850-1849: past1000 1850-2005: historical	1156
MPI-ESM-P	MPI	850-1849: past1000 1850-2005: historical	1156
MRI-CGCM3	MRI	850-1850: past1000 1850-2005: historical	1156

Table A.2: Calibrated localization radii. Localization radii for individual model priors are selected using the radius search and calibration-validation procedure detailed in section A.8.1. Skill metrics are the median values obtained for the mean extratropical MJJA time series relative to BEST for the set of validation periods.

Model	Localization Radius (km)	Correlation	RMSE (°C)	$\sigma$ Ratio	Mean Bias (°C)
BCC	$\infty$	0.69	0.18	1.03	0.05
CCSM4	16500	0.72	0.19	1.18	0.07
CESM	$\infty$	0.72	0.18	1.08	0.06
CSIRO	$\infty$	0.70	0.19	1.18	0.05
F-GOALS	$\infty$	0.70	0.18	1.02	0.07
HadCM3	$\infty$	0.69	0.19	1.18	0.05
IPSL	12750	0.70	0.19	1.19	0.06
MIROC	26375	0.71	0.19	1.18	0.06
MPI	27625	0.69	0.20	1.18	0.06
MRI	$\infty$	0.71	0.17	1.01	0.05

Table A.3: Temperature field reconstructions used to compare spatial patterns of climate response to radiative forcings in this study. We provide a reference for each CFR along with the name used in this study. We also note the maximum size of the proxy network used in each study along with the target temperature fields.

Name	Reference	Network Size	Reconstruction Target
NTREND - DA	This study	54	MJJA
NTREND - PPR	Anchukaitis et al. (2017)	54	MJJA
Guillet 2017	Guillet et al. (2017)	28	Highpass JJA
Zhu 2020	Zhu et al. (2020)	395	JJA
LMR 2.1	Tardif et al. (2019)	544	Annual (Jan. - Dec.)
Neukom (DA)	Neukom et al. (2019)	210	Annual (April - March)

Table A.4: Withheld proxy verification statistics for individual models. Reported skill metrics are the median for all individual proxy comparisons over the 54 leave-one-out assimilations.

Model	Correlation	RMSE	$\sigma$ Ratio	Mean Bias °C
BCC	0.53	0.98	0.42	0.12
CCSM4	0.52	0.98	0.42	0.06
CESM	0.50	1.03	0.35	0.27
CSIRO	0.54	1.01	0.31	0.13
F-GOALS	0.47	1.04	0.34	0.06
HadCM3	0.49	1.03	0.39	0.25
IPSL	0.53	1.00	0.38	0.08
MIROC	0.53	1.01	0.37	0.25
MPI	0.53	0.99	0.39	0.11
MRI	0.55	0.98	0.32	0.16

Table A.5: As in Table A.2, but using the RMSE optimization scheme.

Model	Localization Radius (km)	Correlation	RMSE ( $^{\circ}\text{C}$ )	$\sigma$ Ratio	Mean Bias ( $^{\circ}\text{C}$ )
BCC	18875	0.71	0.17	0.78	0.06
CCSM4	7375	0.71	0.18	0.81	0.07
CESM	15750	0.71	0.18	0.84	0.07
CSIRO	15750	0.70	0.18	0.80	0.06
F-GOALS	19000	0.72	0.18	0.77	0.08
HadCM3	13375	0.70	0.18	0.82	0.06
IPSL	6750	0.70	0.18	0.80	0.07
MIROC	11125	0.71	0.18	0.84	0.07
MPI	10250	0.70	0.18	0.80	0.07
MRI	20250	0.71	0.17	0.78	0.06

Table A.6: Pseudo-proxy localization radii and split-sample validation metrics. As in Table A.2, but using climate model output as the target field.

Target	Prior	Localization Radius (km)	Correlation	RMSE (°C)	$\sigma$ Ratio	Mean Bias (°C)
CESM	CESM	$\infty$	0.73	0.18	0.76	0.02
CESM	MPI	$\infty$	0.72	0.19	0.91	0.02
MPI	CESM	$\infty$	0.74	0.21	0.62	0.09
MPI	MPI	$\infty$	0.75	0.20	0.75	0.07

Table A.7: Skill metrics for pseudo-proxy reconstructions of mean extratropical May-August time series. DA reconstructions use the realistic biased-model, noisy-proxy, time-attrition experimental design. PPR time series and target time series are calculated using only the grid cells for which  $RE > 0$  in each reconstructed time step.

Target Field	Reconstruction Method	Correlation	RMSE ( $^{\circ}\text{C}$ )	$\sigma$ Ratio	Mean Bias ( $^{\circ}\text{C}$ )
CESM	DA, MPI Prior	0.67	0.20	0.84	-0.03
	PPR	0.68	0.25	0.96	0.03
MPI	DA, CESM Prior	0.74	0.41	0.66	0.35
	PPR	0.73	0.46	0.84	0.37

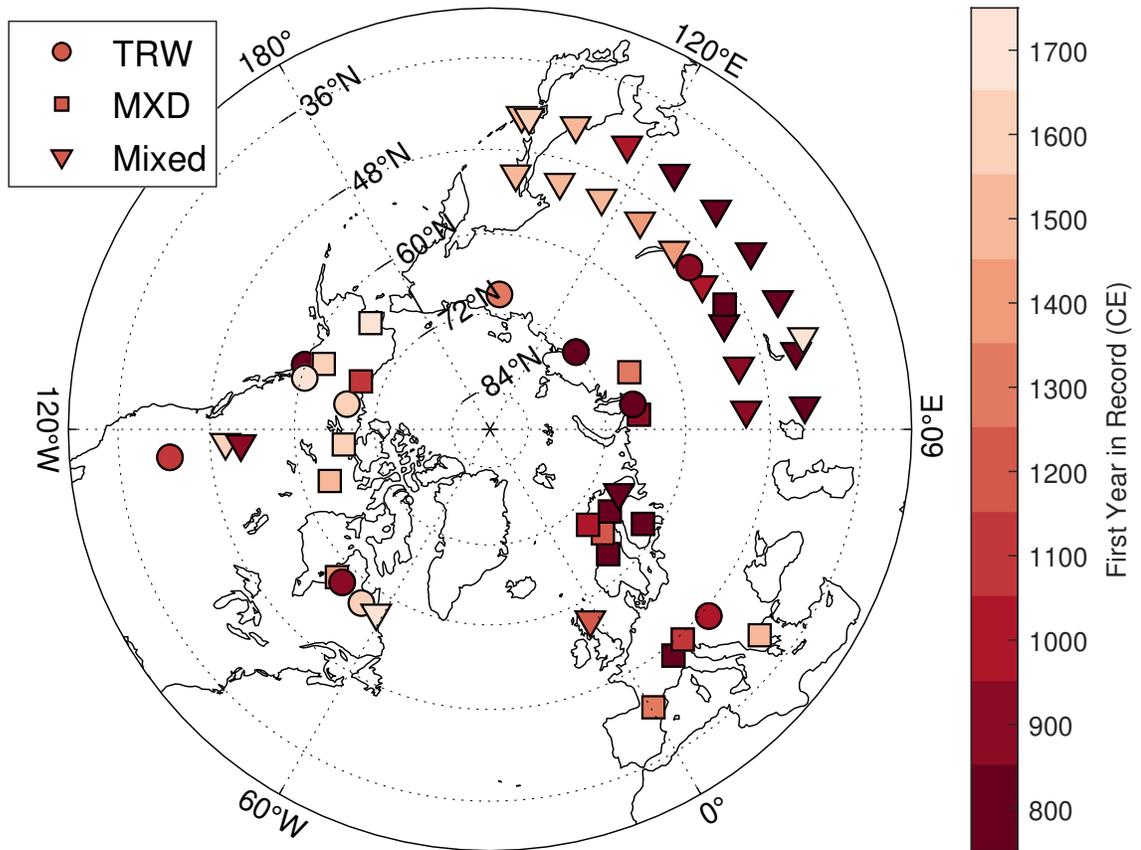


Figure A.1: Locations of the 54 NTREND sites (Wilson et al., 2016). NTREND records were developed using ring-width data (TRW; circles), maximum latewood density (MXD; squares), or a mix of TRW, MXD, and blue intensity (Mixed; triangles). Marker color denotes the century in which each record begins.

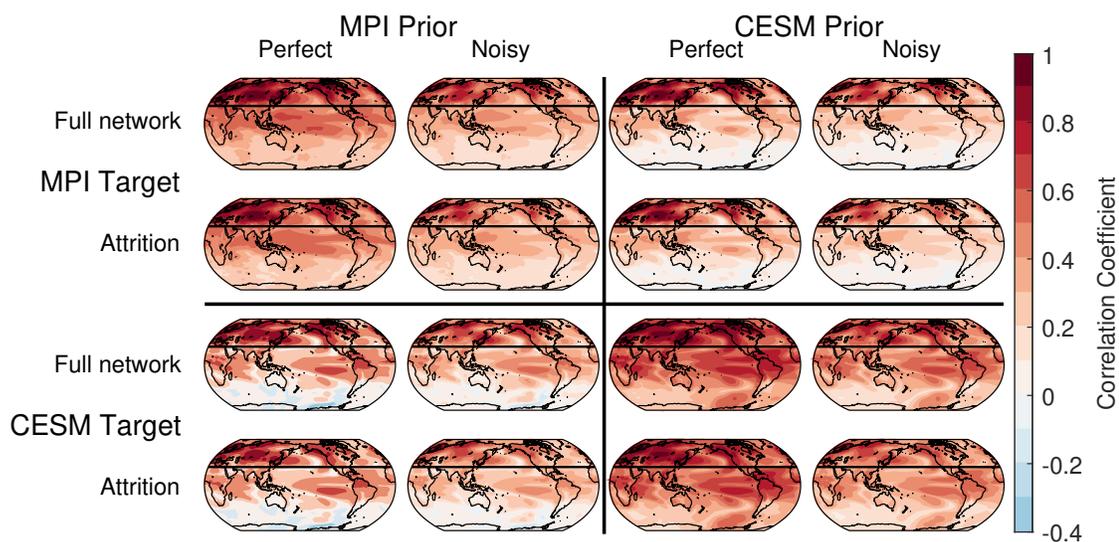


Figure A.2: Local Pearson's correlation coefficients of pseudo-proxy reconstruction temperature anomalies with the target fields. Correlation coefficients are calculated over the period 850-1988 CE. Major rows indicate the model used to generate the target field, and major columns show the model used to build the initial ensemble for each assimilation. Minor rows designate whether the proxy network exhibits no time attrition or realistic time attrition. Minor columns indicate whether reconstructions use perfect or noisy proxies. The top-left and bottom-right quadrants display the perfect-model experiments, while the top-right and bottom-left quadrants show the biased-model cases. The black line in each map indicates 30°N.

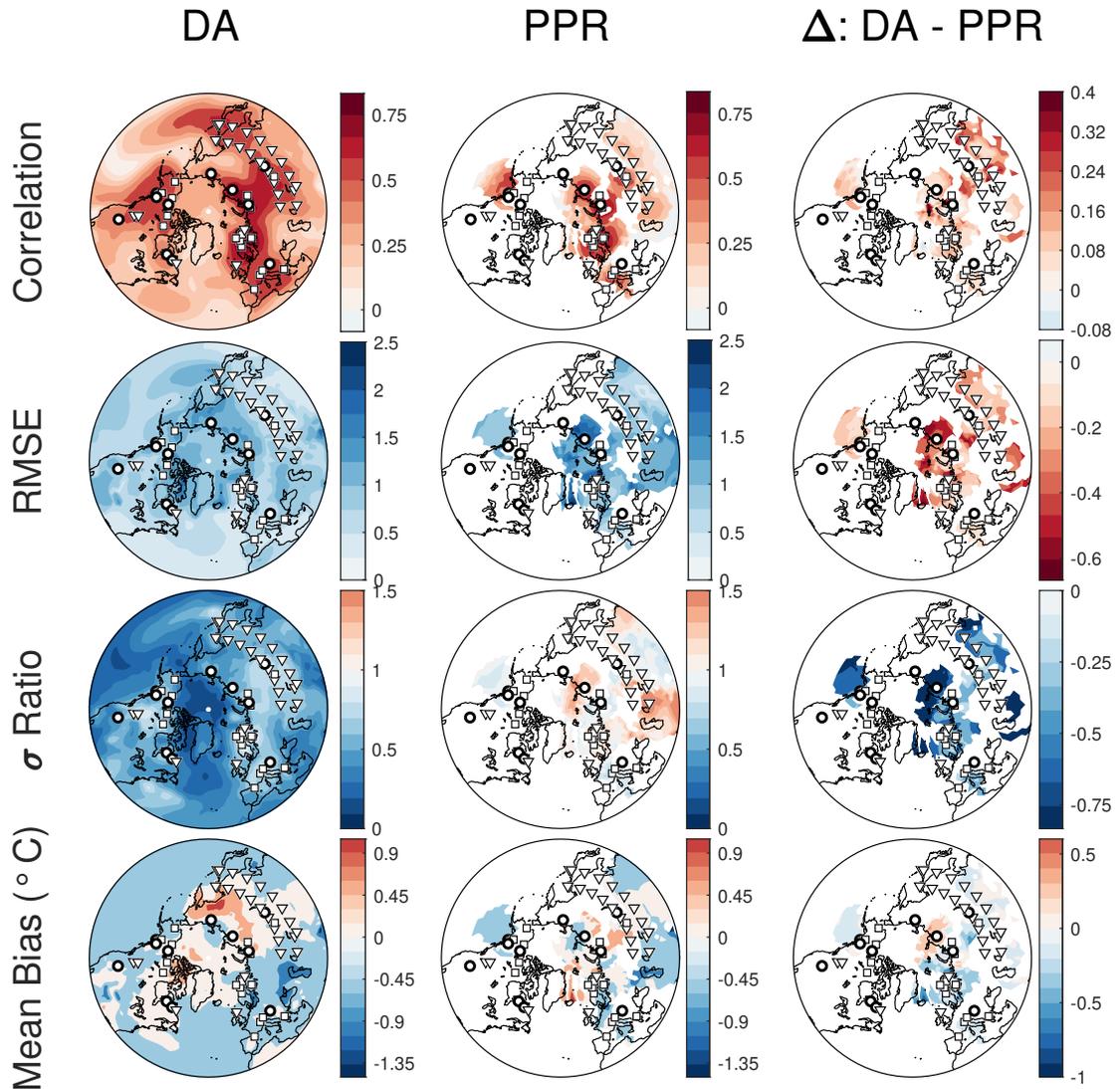


Figure A.3: Pseudo-proxy reconstruction skill for DA (left column), PPR (middle), and a comparison of the two (right). Skill metrics are relative to a CESM target field using noisy proxies and realistic temporal attrition. DA results are for a biased-model MPI prior. All skill metrics are computed over the period 850-1988 CE. In order the rows detail local Pearson's correlation coefficients, RMSE values, temporal standard deviation ( $\sigma$ ) ratios, and mean biases. Comparison plots show DA skill minus PPR skill. The comparison plot of  $\sigma$  ratios only considers grid points where  $\sigma$  is underestimated in both the DA and PPR reconstruction.

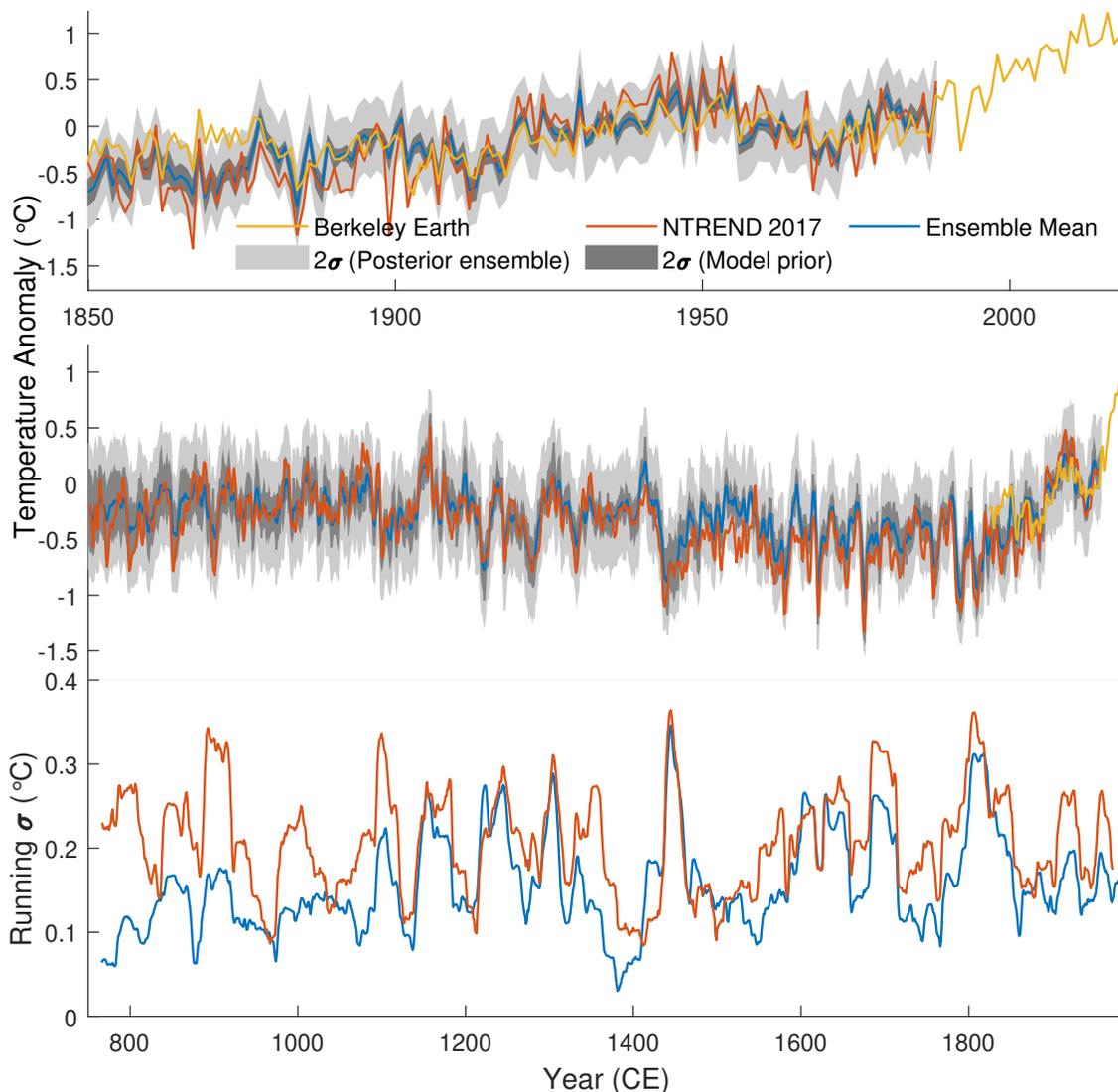


Figure A.4: Extratropical MJJA time series for the multi-model mean reconstruction (blue), Berkeley Earth instrumental records (yellow), and Anchukaitis et al. (2017) (red). We provide two different measures of uncertainty for the DA time series: the average of the  $2\sigma$  posterior ensemble width taken over the 10 reconstruction (light grey), and the  $2\sigma$  width of variability arising from prior model selection (dark grey). Reconstructed temperature anomalies are shown in Celsius for the instrumental era (top), and full reconstruction (middle). A three year moving average has been applied to the time series in the middle panel. The bottom panel displays the 31-year, running standard deviation of the DA ensemble-mean and Anchukaitis et al. (2017) time series.

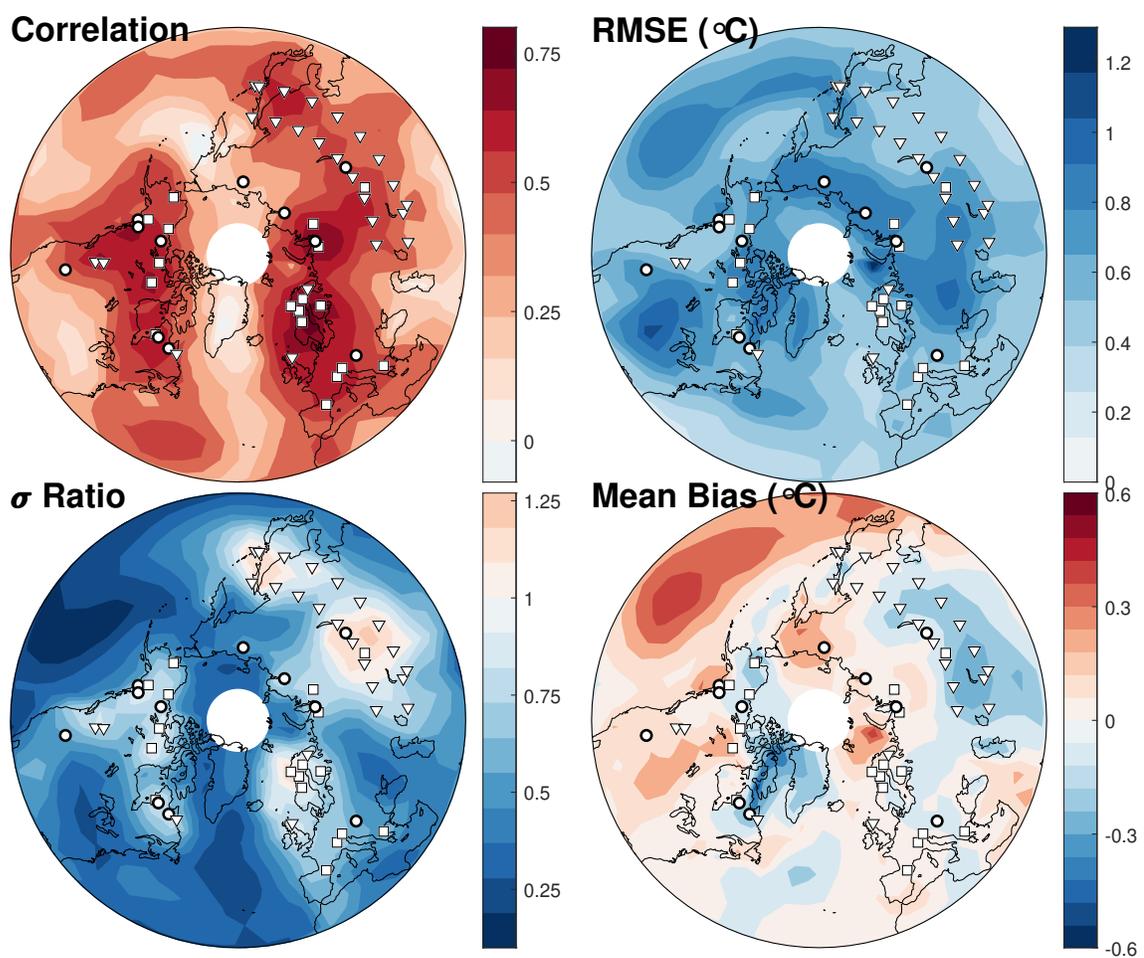


Figure A.5: Spatial skill metrics for the multi-model mean reconstruction. Maps detail Pearson correlation coefficients (top left), RMSE values (top right),  $\sigma$  ratios (bottom left), and mean biases (bottom right) of reconstructed grid point time series relative to the Berkeley Earth instrumental dataset over the period 1901-1988 CE. White markers show the proxy network and marker symbols follow the convention in Figure A.1.

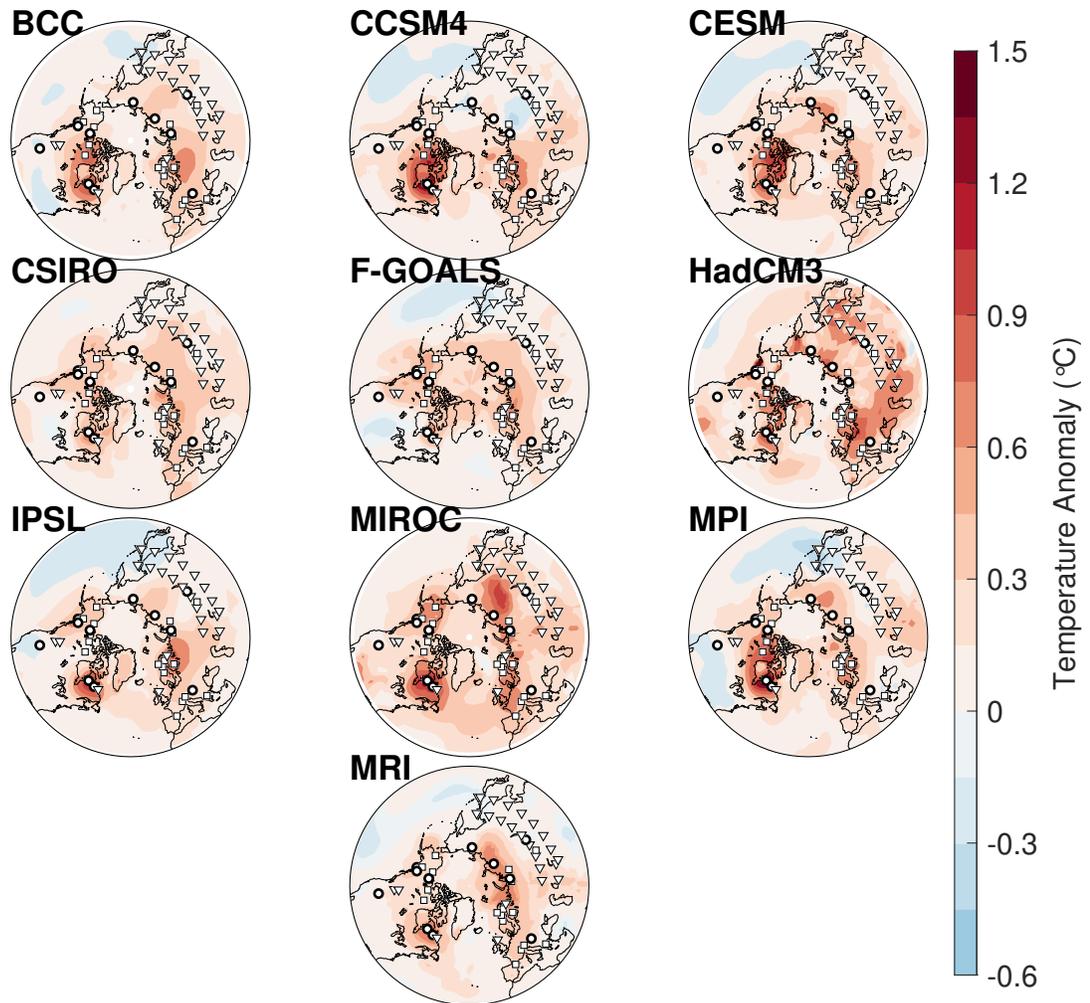


Figure A.6: Reconstructed temperature anomalies (in Celsius) between the MCA (950-1250 CE) and LIA (1450-1850 CE) for the DA reconstructions. Each map shows the results for a particular model prior.

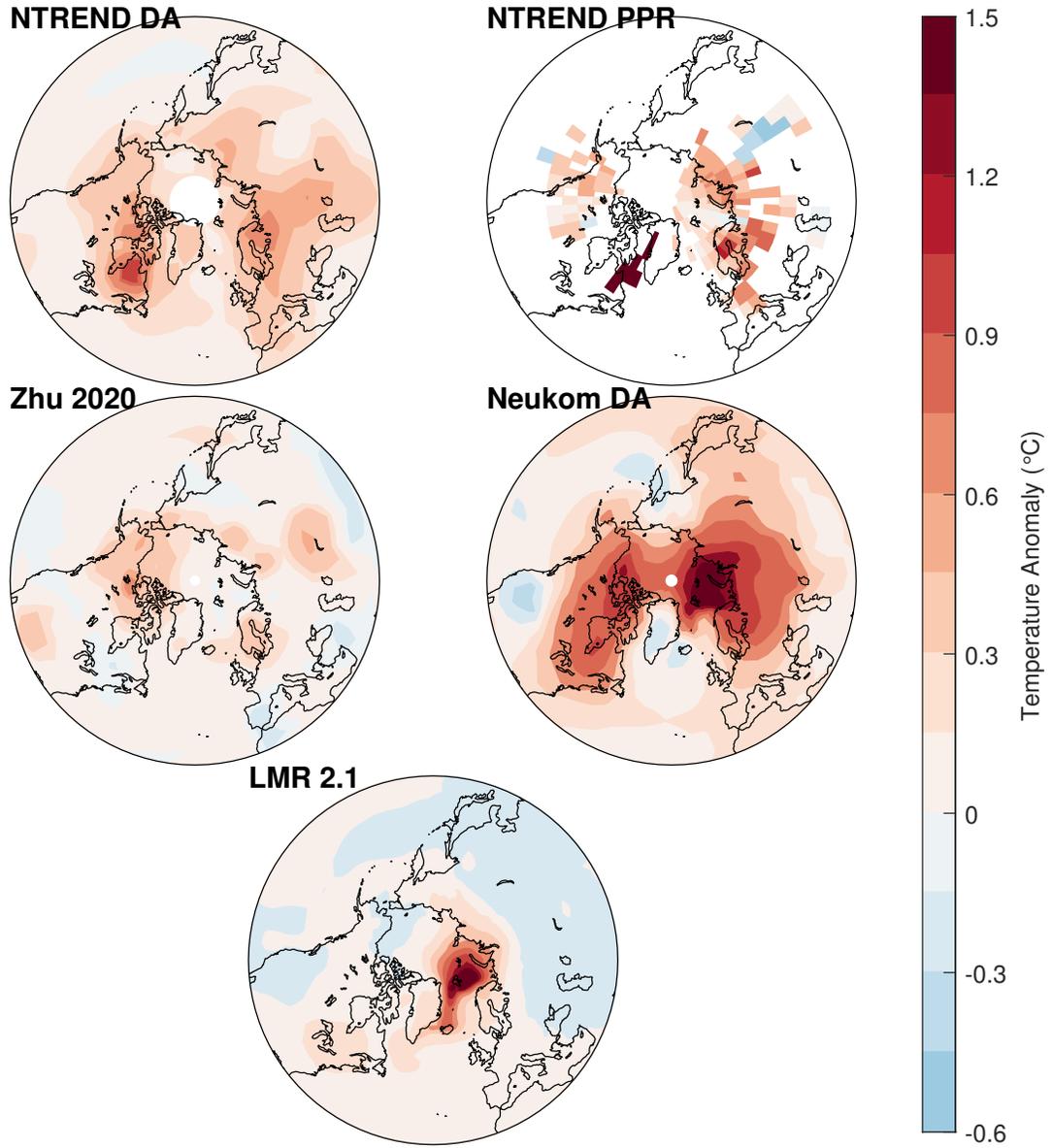


Figure A.7: As in A.6, but for the temperature CFRs summarized in Table A.3.

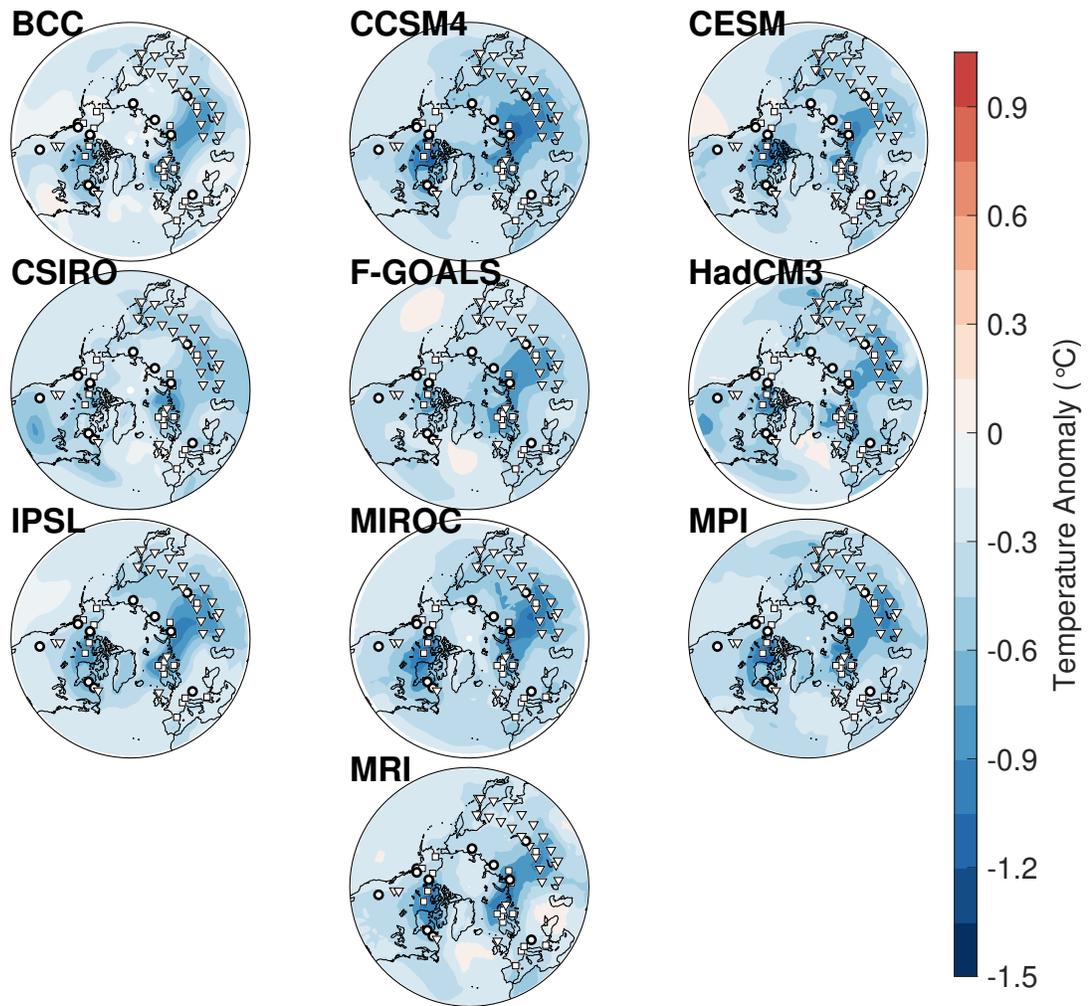


Figure A.8: Composite mean maps of the reconstructed temperature response in years containing a major tropical volcanic event. Events ( $N=20$ ) are selected as tropical eruptions with a global forcing magnitude equal or larger than the 1884 Krakatoa eruption: this set consists of 916, 1108, 1171, 1191, 1230, 1258, 1276, 1286, 1345, 1453, 1458, 1595, 1601, 1641, 1695, 1809, 1815, 1832, 1836, and 1884 CE (Sigl et al., 2015; Anchukaitis et al., 2017). Temperature anomalies (in Celsius) are determined relative to the mean temperature of the five years preceding each volcanic event. Each map shows the results for a particular model prior.

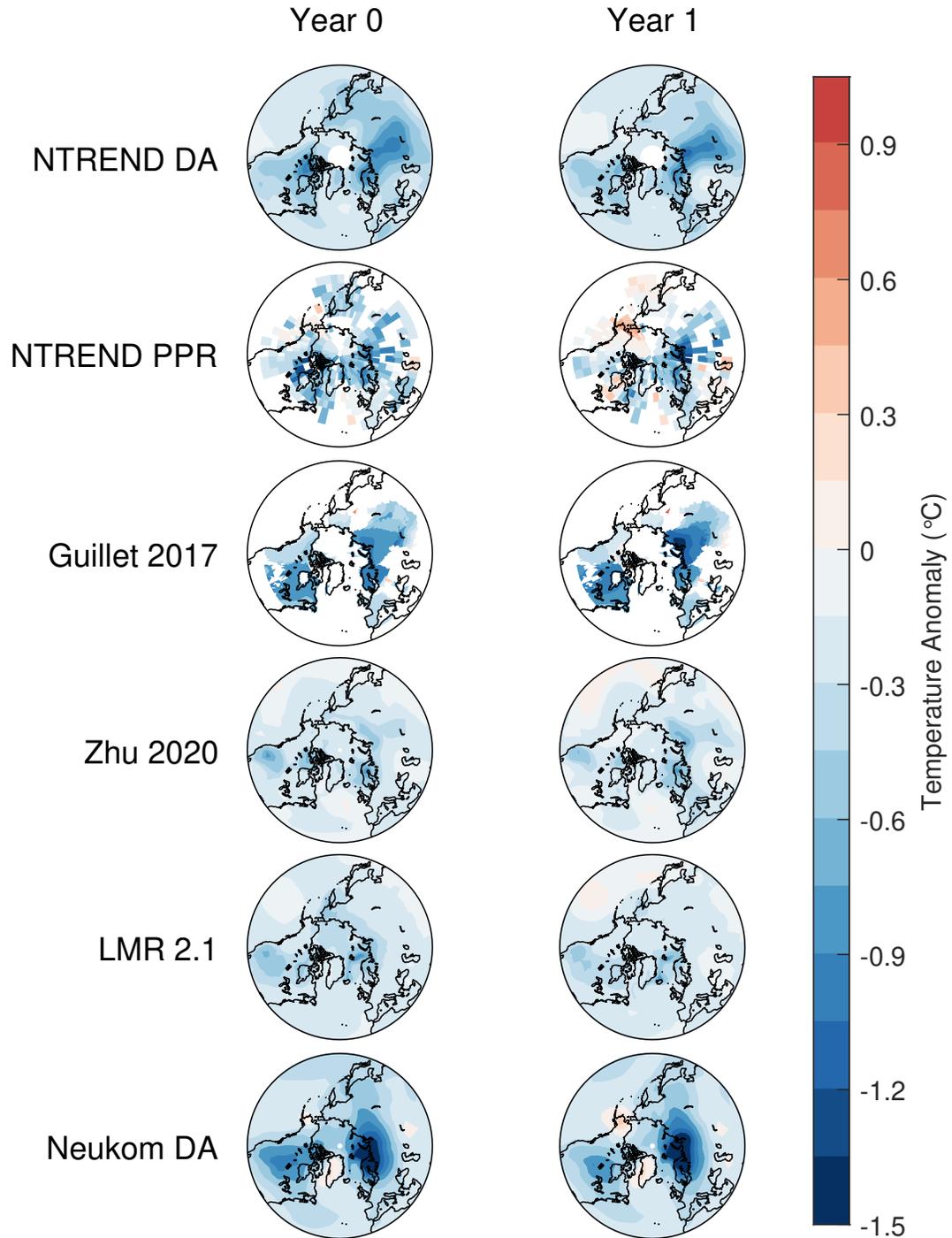


Figure A.9: As in Figure A.8, but for the temperature CFRs summarized in Table A.3 (rows). We only show grid points with reconstructed values for at least 6 eruptions. Maps show the composite mean response in years with a major tropical eruption (left), and in the year following a major eruption (right).

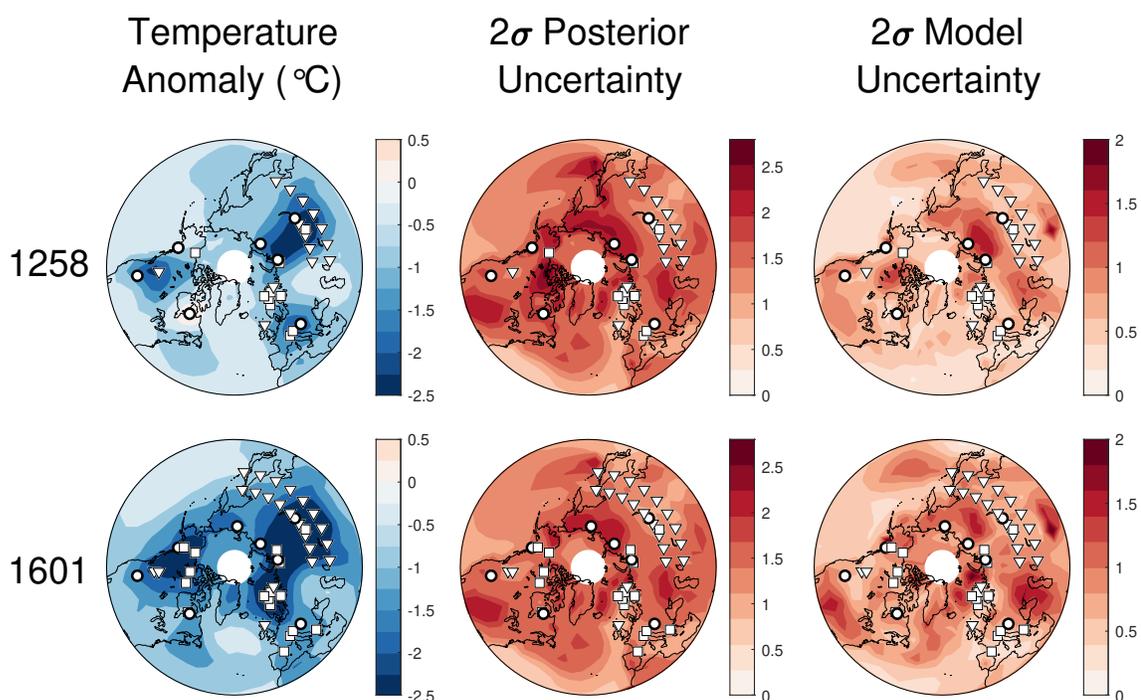


Figure A.10: Spatial characteristics in the year following volcanic eruptions in 1257 (top) and 1600 (bottom) (De Silva and Zielinski, 1998; Lavigne et al., 2013) in the multi-model mean reconstruction. The left column displays temperature anomalies relative to the five preceding years in Celsius. The middle column shows the average 2 $\sigma$  width of the 10 posterior ensembles, and the right column shows the 2 $\sigma$  width of the multi-model ensemble. White markers show the proxy network for each event. Marker symbols follow the convention in Figure A.1.

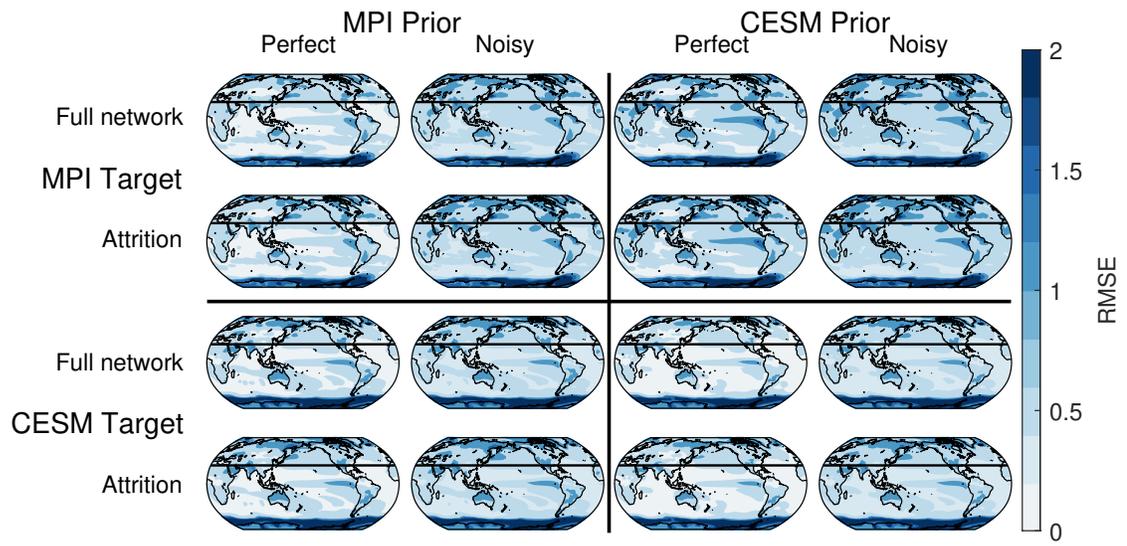


Figure A.11: As in Figure A.2, but for RMSE (°C).

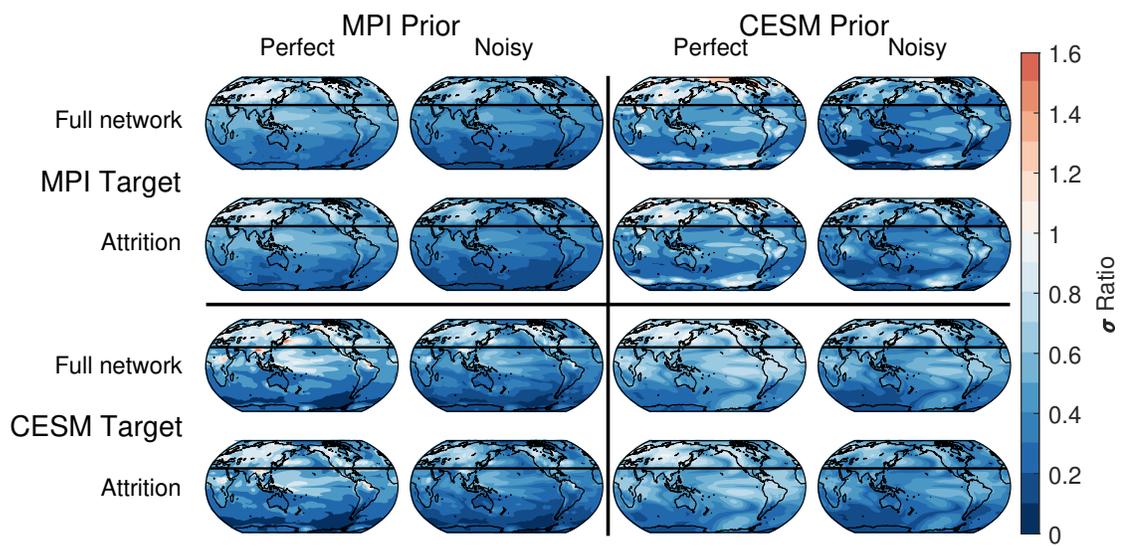


Figure A.12: As in Figure A.2, but for  $\sigma$  ratios.

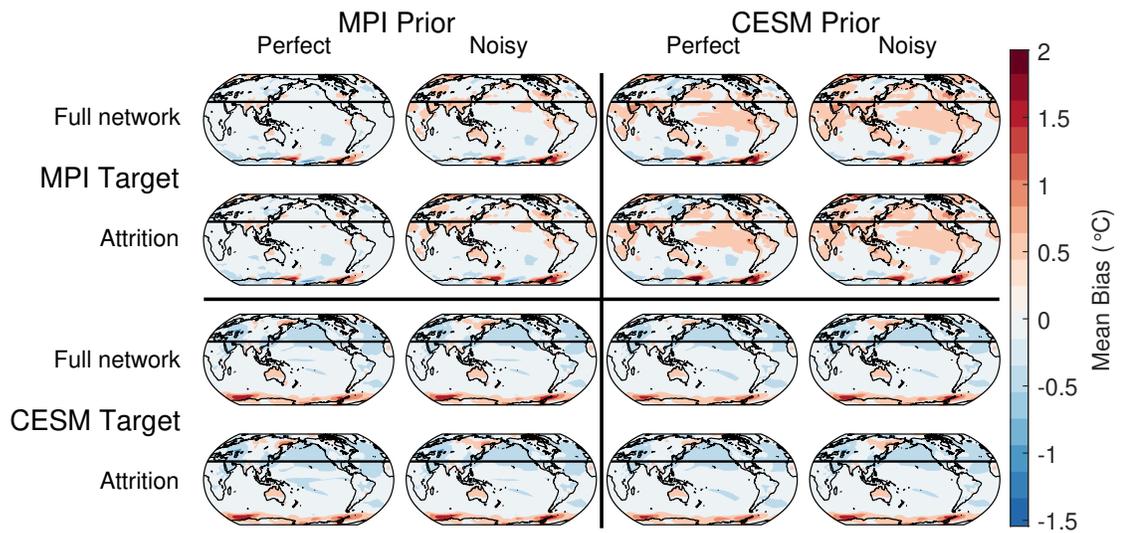


Figure A.13: As in Figure A.2, but for mean biases (°C).

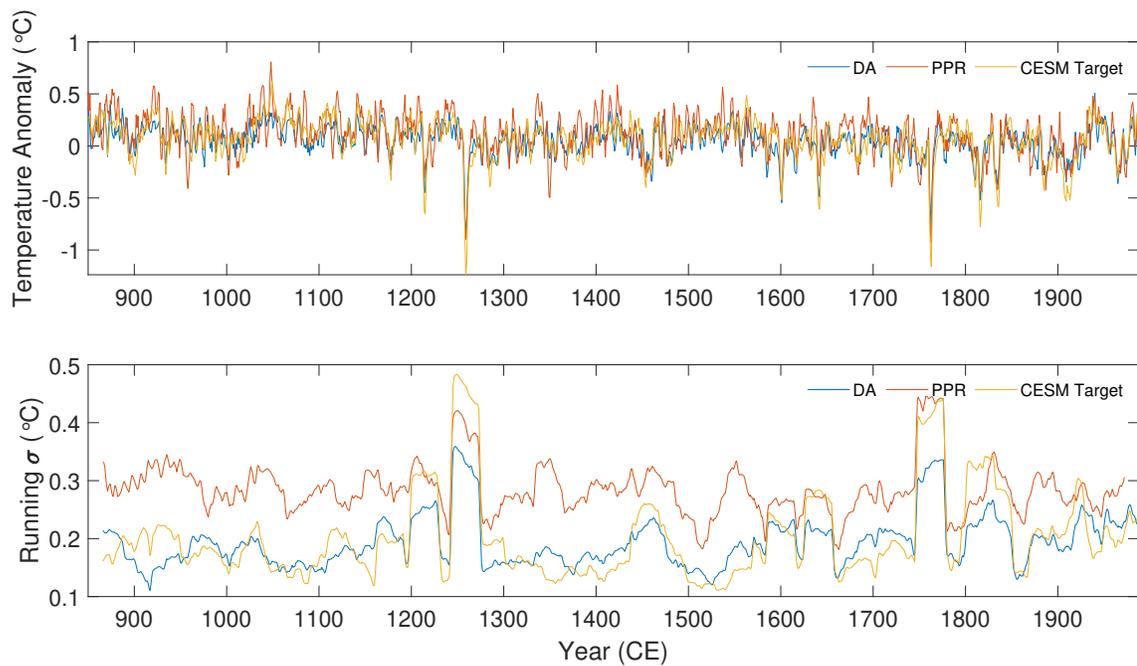


Figure A.14: Extratropical MJJA time series for the pseudo-proxy experiments with a CESM target. Reconstructed temperature anomalies are shown in Celsius (top) for the DA reconstruction (blue) and PPR reconstruction (red) along with the reconstruction target (yellow). The bottom panel displays a 31 year running standard deviation for each time series. A three year moving average has been applied to all time series.

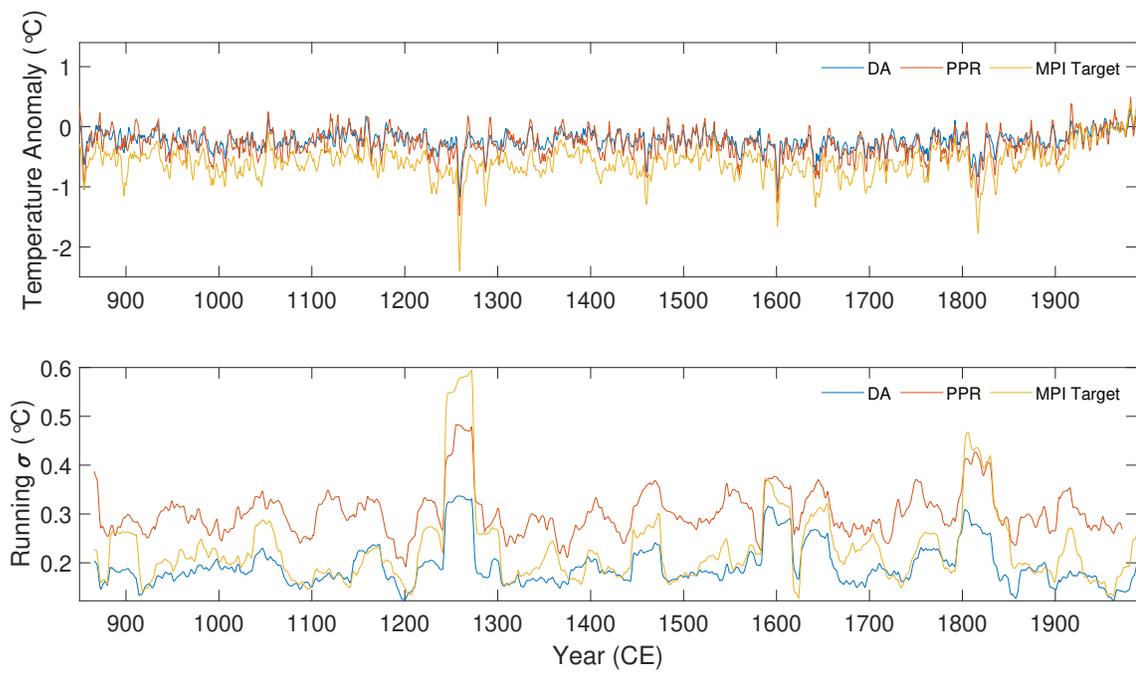


Figure A.15: As in Figure A.14, but for an MPI target.

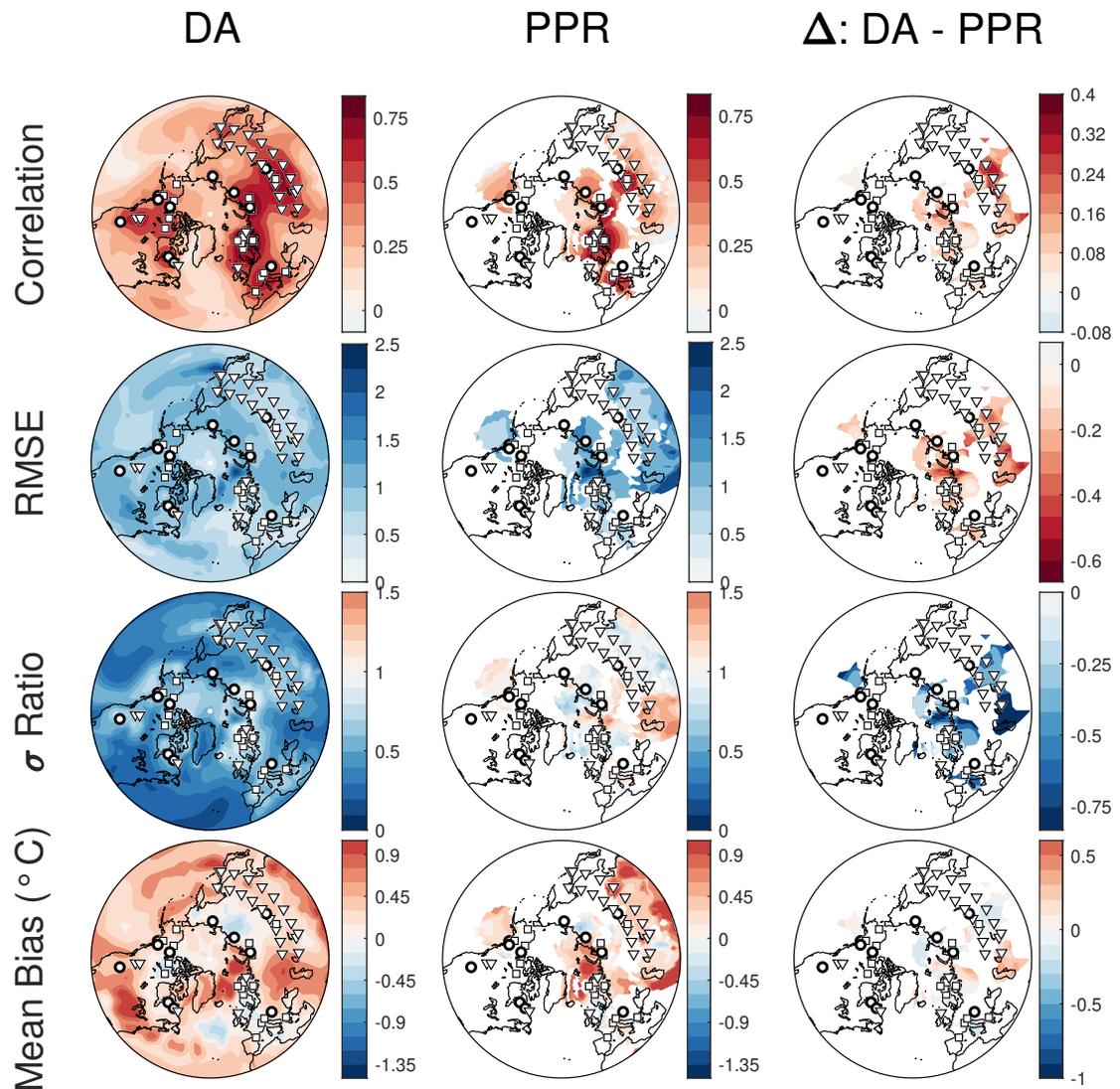


Figure A.16: As in Figure A.3, but for a MPI target field. Here, the DA reconstructions use a CESM prior.

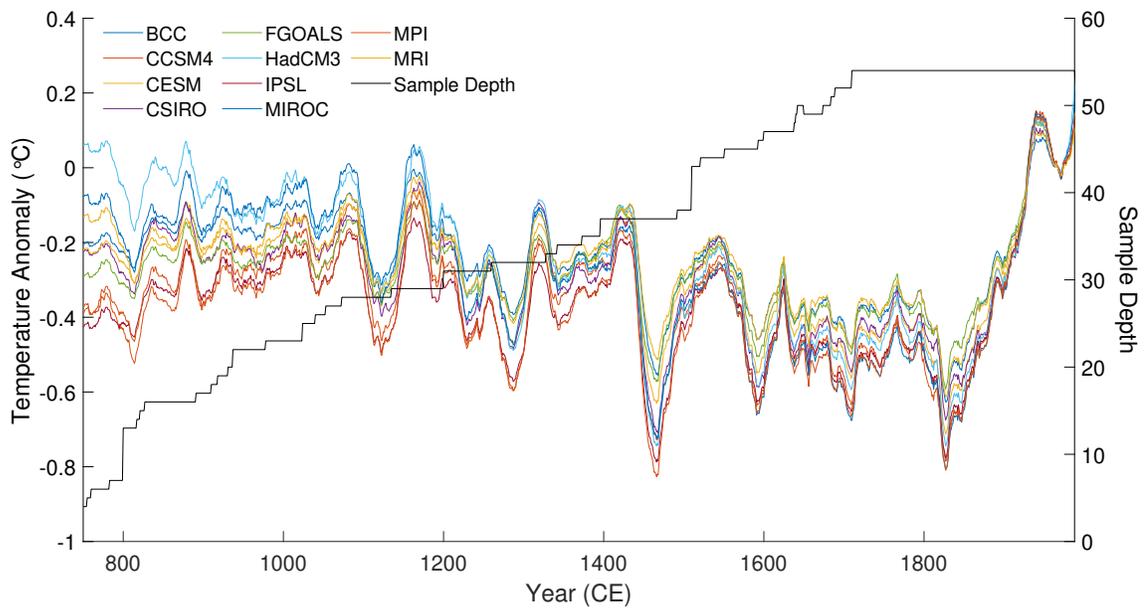


Figure A.17: Extratropical MJJA time series for the individual DA reconstructions. Each time series shows the results for a particular model prior. A 31 year moving average has been applied to each time series.

## APPENDIX B

## Trends and variability in the Southern Annular Mode over the Common Era

King, J., Anchukaitis, K., Allen, K., Vance, T., Hessler, A. Trends and variability in the Southern Annular Mode over the Common Era. *Science Advances* (in review).

## Trends and variability in the Southern Annular Mode over the Common Era

Jonathan King<sup>1,2</sup>, Kevin Anchukaitis<sup>2,3,1</sup>, Kathy Allen<sup>4,5,6</sup>, Tessa Vance<sup>7</sup>,  
Amy Hessl<sup>8</sup>

<sup>1</sup> Department of Geosciences, University of Arizona, Tucson, AZ 85721 USA

<sup>2</sup> Laboratory of Tree-Ring Research, University of Arizona, Tucson, AZ 85721 USA

<sup>3</sup> School of Geography, Development, and Environment, University of Arizona, Tucson,  
AZ 85721 USA

<sup>4</sup> School of Geography, Planning and Spatial Sciences, University of Tasmania, Hobart  
7001

<sup>5</sup> School of Ecosystem and Forest Sciences, University of Melbourne, Richmond, VIC  
Australia 3121

<sup>6</sup> Centre of Excellence for Australian Biodiversity and Heritage, University of NSW,  
Australia

<sup>7</sup> Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies,  
University of Tasmania, Hobart, Australia

<sup>8</sup> Department of Geology and Geography, West Virginia University, Morgantown, WV  
USA

### B.1 Abstract

The Southern Annular Mode (SAM) is the leading mode of atmospheric variability in the extratropical Southern Hemisphere and has wide ranging effects on ecosystems and societies. Despite the SAM's importance, paleoclimate reconstructions disagree on its variability and trends. Here, we use data assimilation to reconstruct the SAM over the last 2000 years using temperature and drought-sensitive climate proxies. Our method does not assume a stationary relationship between proxy records and the SAM over an instrumental calibration period, so our reconstruction is less sensitive to the teleconnection variability that has hindered previous reconstructions. Our approach also allows us to identify critical paleoclimate records and quantify reconstruction uncertainty through time. We find no

evidence for a forced response in SAM variability prior to the 20th century. We also find the modern positive trend is outside the range of the prior 2000 years, but only on multidecadal time scales.

## **B.2 Introduction**

The Southern Annular Mode (SAM) is the leading mode of atmospheric variability in the extratropical Southern Hemisphere and is characterized by a mostly zonally-symmetric mass oscillation with anti-correlated anomalies over the mid-latitudes and Antarctica (Rogers and Van Loon, 1982; Gong and Wang, 1999; Thompson and Wallace, 2000; Marshall, 2003). The SAM's phases capture the strength and position of the mid-latitude westerly winds and the subtropical jet, such that positive phases indicate a poleward shift and intensification of the circumpolar westerly belt, while negative phases indicate an equatorward shift and weakening of the westerly winds. Variability in the SAM's phases ultimately results from changes to the meridional pressure gradient across the mid-latitudes of the Southern Hemisphere, which are driven by jet variability, connections with the stratosphere, and ENSO (see Fogt and Marshall, 2020, and references therein). Because of its influence on the polar jet, the SAM has wide ranging climate effects across the Southern Hemisphere. Positive phases of the SAM are linked to cooling over Australia and central Antarctica, as well as warming over the Antarctic Peninsula and southern South America (Thompson and Solomon, 2002; Kwok and Comiso, 2002; Gillett et al., 2006; Gupta and England, 2006; Marshall et al., 2006; Hendon et al., 2007; Van Lipzig et al., 2008). The SAM also influences regional hydroclimates through its effects on Southern Hemisphere storm tracks. Effects of the positive phase include drying over southern South America, western South Africa, southern Australia, and New Zealand, as well as increased precipitation over central and eastern Australia and southeastern South America (Silvestri and Vera, 2003, 2009; Cai and Cowan, 2006; Reason and Rouault, 2005; Gillett et al., 2006; Hendon et al., 2007; Kidston et al., 2009), and increased wild-fire activity in South America and south-east Australia (Holz and Veblen, 2011; Mariani and Fletcher, 2016; Holz et al., 2017; Abram et al., 2021; Lim et al., 2021). The SAM

is also linked to changes in sea ice distribution (Hall and Visbeck, 2002; Gupta and England, 2006; Stammerjohn et al., 2008; Simpkins et al., 2012; Kohyama and Hartmann, 2016), and ocean-atmosphere carbon exchange (Lovenduski et al., 2007, 2015; Sallée et al., 2010). Understanding SAM variability is therefore important for both societies and ecosystems throughout the Southern Hemisphere, particularly in sub-tropical-temperate regions projected to experience a future drying climate.

Since the 1950s, the SAM has exhibited a trend toward a more positive state (Thompson et al., 2000; Marshall, 2003; Fogt and Marshall, 2020), which is attributed to stratospheric ozone depletion and rising concentrations of atmospheric CO<sub>2</sub> (Arblaster and Meehl, 2006; Thompson et al., 2011; Polvani et al., 2011; England et al., 2016; Jones et al., 2016). This positive trend has potentially contributed to severe droughts, including the Day Zero Cape Town drought (Sousa et al., 2018) and Millennium Drought in Australia (Verdon-Kidd and Kiem, 2009; Cai et al., 2011), as well as increased fire activity (Holz and Veblen, 2011; Mariani and Fletcher, 2016; Holz et al., 2017). Given these impacts, it is important to place the SAM's recent behavior in a long-term perspective and assess the relative influence of anthropogenic forcing and natural climate variability. In the context of multi-decadal trends, reconstructions spanning multiple centuries are necessary to resolve forced responses from the SAM's internal variability. Instrumental records of the SAM only extend through the mid-1900s and longer reanalysis-derived indices show low correlations with one another and differences in variability prior to the 1950s (Barrucand et al., 2018; Fogt and Marshall, 2020), so characterizing the SAM's long-term behavior requires paleoclimate reconstructions derived from natural climate archives.

There are several existing multi-century SAM reconstructions: Villalba et al. (2012), Abram et al. (2014), and Dätwyler et al. (2018) (henceforth, V12, A14, and D18), but they show limited agreement prior to the 1850s (Hessl et al., 2017; Dätwyler et al., 2018). Most indicate a negative phase in the SAM during the late 1400s, but both trends and decadal-scale variability show large discrepancies aside from this feature. There are several potential reasons for these differences. Firstly, all three reconstructions rely on the calibration of proxy records directly with an instrumental SAM index. This implic-

itly makes two important assumptions for each reconstruction: first, that the relationship of proxy records to local climate variables is stationary over time; and second, that the SAM's teleconnections with local climate variables are stationary and well-represented by the instrumental record. While the first is reasonable and a necessary assumption of most paleoclimate analyses, multiple studies cast doubt on this second point, and regional complexity in the climate response to specific SAM phases further decreases the likelihood of this assumption holding. For instance, even over the instrumental period, SAM exhibits non-stationary connections with precipitation and temperature anomalies in southern South America, Australasia, and the Antarctic Peninsula (Silvestri and Vera, 2009; Gallant et al., 2013), and many of the proxy records in existing SAM reconstructions come from these areas (Hessl et al., 2017). Evolving concentrations of greenhouse gases, stratospheric ozone, connections with ENSO, and stochastic climate variability can also affect the SAM's influence on regional climates over multi-decadal time scales (Thompson et al., 2000, 2011; Brönnimann et al., 2017; Yun and Timmermann, 2018). Pseudo-proxy experiments have also shown that non-stationary teleconnections cause reconstruction skill to vary widely with the selection of different calibration windows (Huiskamp and McGregor, 2021). This effect is particularly pronounced for proxy networks with fewer than 20 sites, which is common in the early portions of SAM reconstructions. To mitigate such effects, D18 explicitly screened for stationarity in their reconstruction, although this required calibration with a longer and therefore less reliable observational record (Jones et al., 2009b).

Differences between SAM reconstructions may also result from the selection of different reconstruction targets and proxy networks. For example, A14 targets an annual SAM index, whereas V12 and the "Best Reconstruction" of D18 target an austral summer (DJF) SAM index. D18 found that annual reconstructions were much more sensitive to the selection of proxy sites and calibration windows and they conclude that annual products may exhibit increased sensitivity to non-stationary teleconnections, which may partly explain the differences between the reconstructions. Additionally, each index has been reconstructed using a different proxy network with a different geographic extent. A14 targets the Drake Passage sector, using a mix of terrestrial proxy types from south-

ern South America as well as Antarctic ice cores. In comparison, V12 targets the Pacific sector, using a network of tree-ring chronologies from South America and New Zealand. D18 uses the most spatially extensive network, including tree-ring records (Villalba et al., 2012), Antarctic ice cores, PAGES2k South American proxies (PAGES2k Consortium, 2017), and coral records from the tropical Pacific (Tierney et al., 2015). Furthermore, A14 utilizes a temperature-sensitive proxy network, while V12 and D18 leverage both temperature and hydroclimate-sensitive proxies. Given the variability of the SAM's teleconnections on regional scales (Silvestri and Vera, 2009; Gallant et al., 2013), and the climate sensitivities of different proxy types (Jones et al., 2009b), these variations in proxy-network design may further help explain reconstruction differences. It is often difficult to assess the influence and contribution of individual proxy records in multiproxy reconstructions, so the cause of any reconstructed index's behavior are often unclear. This is particularly relevant in the period prior to 1400 CE, when the sparsity of proxy networks leaves the reconstructions vulnerable to the dominant influence of just a few records. Ultimately, as a consequence of these uncertainties and the differences in existing reconstructions, the evolution of the SAM over the Common Era and its response to external forcing remains poorly constrained (Hessl et al., 2017; Gulev et al., 2021).

To address these uncertainties, here we reconstruct the austral summer (DJF) SAM index over the Common Era at annual resolution using offline paleoclimate data assimilation (DA). DA is a recently developed reconstruction technique that integrates climate proxy records with the dynamical behavior captured by climate models (Steiger et al., 2014; Hakim et al., 2016). In brief, DA uses forward or proxy-system models (Evans et al., 2013) to translate climate model states into the same dimensions or 'space' as a collection of climate proxy records. This allows direct comparison of the model output with the proxy records. The climate model states are then updated to more closely match the proxy records, and a model-derived estimate of climate system covariance is used to propagate the update to reconstruction targets, such as the SAM. DA has recently been used to reconstruct surface air temperature anomalies (Goosse et al., 2012; Tardif et al., 2019; Neukom et al., 2019; King et al., 2021), geopotential height fields (Tardif et al., 2019), the response to volcanic eruptions (Zhu et al., 2020, 2022), sea ice extent

(Dalaiden et al., 2021), sea surface temperatures (Tierney et al., 2020), and hydroclimate variables (Steiger et al., 2018). In this study, we assimilate the PAGES2k temperature-sensitive proxy network (PAGES2k Consortium, 2017), the South American Drought Atlas (Morales et al., 2020, SADA), and the Australia-New Zealand Drought Atlas (Palmer et al., 2015, ANZDA) (Figure B.1) using a suite of last millennium general-circulation climate models to reconstruct the austral summer SAM index over the last 2,000 years.

In the context of SAM reconstructions, DA offers several additional advantages relative to traditional methods. Firstly, our method does not calibrate proxy records against an instrumental SAM index directly; instead, we calibrate proxy forward models using local climate variables, like temperature and precipitation, near the proxy sites. Consequently, our calibration does not assume stationary SAM teleconnections and only requires the stability of proxy relationships to their local climate. Additionally, we estimate covariance between proxies and the SAM using thousands of years of climate model output. As a result of this, our proxy-SAM relationships are not sensitive to potentially anomalous decadal- or centennial-scale variations in the SAM's behavior. Furthermore, DA is amenable to the use of a range of proxy types as well as gridded climate records with spatial autocorrelation, and we leverage this to incorporate the two existing tree-ring based drought atlases into our reconstruction. Previous work indicates that SAM reconstructions using hydroclimate-sensitive sites are more skillful than those using strictly temperature-sensitive proxy networks (Huiskamp and McGregor, 2021). Each drought atlas provides extensive coverage for at least the last five centuries and each incorporates over 150 tree-ring records. They therefore represent a significant source of hydroclimate information available for our reconstruction.

Finally, our DA method allows us to incorporate an optimal sensor analysis (Comboul et al., 2015) as part of the final reconstruction. Traditionally, optimal sensor analyses have been used to identify ideal regions for future proxy development (Bradley, 1996; Evans et al., 1998; Mauger et al., 2013; Comboul et al., 2015); however, they can also be applied within a DA framework to quantitatively assess the power of different proxy sites as the overall network evolves through time. We use this to identify the proxy sites that are most likely to drive the reconstruction in each time step, which helps characterize the

reconstruction's overall behavior. This information is particularly useful in the early part of this Common Era reconstruction, when the sparse network size can give high weights to a limited number of records.

### B.3 Results

We assess the skill of our SAM reconstruction relative to the Marshall and Fogt indices (Table B.1; Figure B.2a). Before comparing time series, we first normalize the Fogt index and our reconstruction to the Marshall index, such that the mean and variance of the detrended normalized time series match those of the detrended Marshall index over the period 1958-2000 CE. This places all series in the same unit space while preserving differences in the instrumental trend. Correlation with the Marshall index (1958-2000 CE) is  $r = 0.72$  ( $p \ll 0.001$ ), which is comparable to that reported for A14 ( $r = 0.75$ ,  $p \ll 0.001$ ). With respect to the 20th century Fogt index, our reconstruction correlates at  $r = 0.65$ ,  $p \ll 0.001$ , somewhat higher than A14 ( $r = 0.51$ ,  $p \ll 0.001$ ). Our RMSE values with the Marshall index (1.45) are similar to, albeit slightly higher than to those reported by D18 (1.32). We emphasize that our reconstruction is not calibrated directly to the SAM index, so the agreement with the Marshall and Fogt indices is not built-in to our reconstruction method and thus represents a more independent skill metric.

We next characterize the reconstruction's behavior over the last two millennia (Figure B.2b). The reconstruction exhibits minimal evidence for trends over most of the first millennium of the Common Era, although the third and seventh centuries are both marked by increased multidecadal variability as the SAM alternates between negative and positive phases. A more strongly negative anomaly in the early 1000s is followed by a notable 100-year positive trend that concludes with the most positive anomalies outside of the instrumental era. The SAM persists in a positive state until the late 1400s, when it abruptly decreases to strongly negative values. After this event, the index returns to near-zero mean anomalies. It has a peak in the mid-1700s and begins exhibiting a positive trend in the early 1800s. This trend intensifies in the later half of the 20th century, and the reconstruction ends with the most positive SAM anomalies observed during the Common

Era.

Reconstruction uncertainty ranges from  $\pm 4.5$  anomaly units in the early reconstruction to less than 2.3 after 1500 CE (Figure B.2b). We note that, because we use a stationary prior, the reconstruction years are treated as fully independent of one another. While this is common in many reconstruction techniques, it does not represent the reality of the SAM, which exhibits persistence on interannual time scales due to potential connections with the stratosphere (Lim et al., 2018; Byrne et al., 2019), tropical variability (L'Heureux and Thompson, 2006; Wilson et al., 2016), and external forcing (Gillett et al., 2013; Jones et al., 2016). One consequence of this is that the uncertainty estimates shown here likely overestimate the true reconstruction uncertainty. Overall, uncertainty decreases as the reconstruction approaches the present day, a result of the increasing size of the proxy network (Figure B.2c).

We use our optimal sensor framework to identify which proxies are most responsible for reducing reconstruction uncertainty over time (Figure B.3). A proxy's ability to reduce uncertainty corresponds to its influence on the reconstruction, so this analysis also allows us to identify which proxies most strongly influence the reconstruction at a given point in time. The first 900 years of the reconstruction are most strongly affected by the Mt. Read (Tasmania) tree-ring record with additional support from the Plateau Remote, WDC06A, and WDC05A ice cores. At 900 CE, the Oroko (New Zealand) tree ring chronology joins the network and supplants Mt. Read as the most influential record. Two large decreases in reconstruction uncertainty occur in 1400 and 1500 CE, which correspond to the addition of the SADA and ANZDA, respectively.

Examining the reconstruction's response to external forcing, we find no coherence with the solar forcing series and no significant common response to major volcanic eruptions (Figure B.4). By contrast, the reconstruction exhibits significant positive trends in the latter half of the twentieth century. However, these modern trends are only significant on time-scales greater than approximately 40 years; trends over shorter time scales fall within the reconstructed range of trends from natural variability. Examining the Marshall Index, we similarly find that trends shorter than about 35 years are within the reconstructed range of natural variability, but that trends longer than about 35 years fall outside

this range. The Marshall Index exhibits its most positive, significant trends for intervals centered on the early 1980s, consistent with literature linking the SAM's modern positive trend to anthropogenic aerosols and the Antarctic ozone hole Thompson and Solomon (2002); Gillett and Thompson (2003); Polvani et al. (2011); Thompson et al. (2011); England et al. (2016); Jones et al. (2016); Fogt et al. (2017). Although this period is near the end of our reconstruction and less well resolved than preceding decades, we note that the reconstruction similarly exhibits strongly-positive, significant trends centered on the early 1980s. Here we have quantified natural variability using the period 1500-1900 CE, the years including both drought atlases. If we instead use the period of the full reconstruction (1-1900 CE), the tests become more stringent. Significant trend in the reconstruction is limited to the last 55-80 year interval, and Marshall Index trends are only significant when containing the interval 1964-2000 CE. We also experiment with using the early portion of the reconstruction (1-899 CE) to quantify natural variability and find these results are similar to those using the full reconstruction period (Figure B.6).

#### **B.4 Discussion**

Our reconstruction suggests that the SAM is dominated by internal variability at least throughout the pre-industrial Common Era. This finding is in agreement with D18, who likewise found minimal influence of solar and volcanic forcing on their reconstruction. Volcanic signals have likewise been a challenge to detect in Southern Hemisphere temperature reconstructions (PAGES 2k-PMIP3 group, 2015). Some studies have proposed that that an observed relationship between SAM and ENSO (Abram et al., 2014; Ding et al., 2012; Fogt et al., 2011; Fogt and Bromwich, 2006; Wilson et al., 2016) could provide a pathway for solar forcing (Emile-Geay et al., 2013) to influence the SAM (Hessl et al., 2017; Wright et al., 2021); however, our results do not support this mechanism during the the Common Era.

In contrast, our analysis indicates that the most recent multi-decadal trend is outside the range of natural variability and reflects the SAM's response to anthropogenic forcing. We emphasize that this modern trend is only significant for intervals longer than about

40 years when assessed against the 1500-1900 CE period, or intervals of about 55 years when considering the full Common Era. Shorter trend periods remain within the range of natural variability, even for the most recent intervals. The significance of the modern positive trend therefore reflects its anomalous persistence, rather than the amplitude of its decadal-scale variation alone. The significance of these longer trends emphasizes the importance of the paleoclimate record, particularly given the uncertainties in instrumental SAM records prior to the late twentieth century (Ho et al., 2012; Barrucand et al., 2018). We also note that the modern positive trend is only outside of the range of natural variability for trends spanning the years from about 1940-2000 CE. Trends are generally not significant during the early 1900s, and are even negative for the 50 year period centered on the 1930s. These results help establish the onset of the modern positive trend at around 1940 CE. This timing coincides with increasing emissions of ozone-depleting substances and greenhouse gasses, and is consistent with literature attributing the modern trend to stratospheric ozone depletion and rising levels of atmospheric CO<sub>2</sub> (Arblaster and Meehl, 2006; Thompson et al., 2011; Polvani et al., 2011; England et al., 2016; Jones et al., 2016).

We next compare our reconstruction with the V12, A14, and D18 products (Figure B.5). We normalize the mean and variance of each index over the period 1400-1850 CE to allow comparison of the series in the same unit space. We select the year 1400 CE because it is the first year with values for all four reconstructions and we end the normalization in 1850 CE to limit the sensitivity of our comparison to differing representations of the post-industrial trend. All four indices agree on the existence of a strong positive trend during the late twentieth century; however, all show limited coherence with one another prior to about 1850 CE, as noted in previous studies (Hessl et al., 2017; Dätwyler et al., 2018). The limited agreement of these reconstructions reduces confidence in the significance of modern trends (Gulev et al., 2021), and the causes of these discrepancies include differing seasonal expressions, different proxy networks, and the relative weights of proxies within those networks. Additionally, V12, A14, and D18 all rely on calibration with the instrumental SAM index, which can cause uncertainty when there is non-stationarity in the teleconnection of local climate with the SAM. Ultimately, our reconstruction does

not solve the problem of differing reconstructions and similarly shows limited agreement with all of V12, A14, and D18. However, our assimilation does not rely on calibration with the SAM index, and offers a potential improvement by reducing uncertainty from non-stationary teleconnections.

An additional advantage of our reconstruction is the transparency provided by the optimal sensor's assessment of the relative weights and influence of proxy records in our network. In general, we find that our reconstruction is most strongly influenced by the two drought atlases, followed by the Mt. Read (Tasmania), Oroko (New Zealand), and Pink Pine (New Zealand) tree ring chronologies, and also the Plateau Remote, Siple Station, WDC06A, and WDC06B ice cores. We note here that a minor change in reconstruction uncertainty does not imply that a proxy has a weak effect on the reconstruction, because highly influential proxies from the same location may present redundant climate signals. For example, the Pink Pine chronology is the third most potentially influential PAGES2k record (Figure B.3f), but has a relatively small effect on reconstruction uncertainty when added to the network in 1457 CE (Figure B.3c). This is because much of the Pink Pine climate signal is already represented by the nearby Oroko site. However, such redundant sites are valuable because they make the reconstruction less sensitive to non-climatic noise from a single highly-influential proxy record. In the case of Pink Pine and Oroko, spreading the southern New Zealand climate signal over two influential records allows either site to partially correct for non-climatic noise in the other. A proxy's potential influence reflects both its covariance with the SAM and the ability of our proxy estimates to accurately estimate the record. Ultimately, assuming our estimates of climate covariance are accurate, the influential sites are those most likely to contribute skill to the reconstruction.

Overall, we find that tree-ring chronologies from Tasmania and New Zealand, the West Antarctic ice cores, and the drought atlas locations in Tasmania, southern New Zealand, the eastern edge of Australia, and southeast South America all have the greatest potential for reconstructing SAM (Figure B.3ab). This suggests that additional proxy development in these regions, or extensions of shorter existing records such as the Oroko and Pink Pine tree-ring chronologies or the Siple Station ice core, would be valuable for

improving the skill of future SAM reconstructions. However, we caution that location alone is not sufficient for proxy utility and that future proxy development must demonstrate a robust sensitivity to local climate that connects them to the SAM. We also note that, in our optimal sensor framework, a proxy's potential influence is a function of (1) the accuracy of our forward (proxy system) models, and (2) the covariance of the resulting proxy estimates with the SAM in the climate models. As a result, our analysis may currently undervalue proxies from regions with limited climate model agreement, and future improvements in both climate and proxy system models may allow paleoclimate data from other regions to contribute to skillful reconstructions of the SAM.

#### **B.4.1 Caveats and Limitations**

Our DA method does not require a calibration with the instrumental SAM, which limits its sensitivity to non-stationarity in the SAM during the instrumental era. However, the trade-off is the influence of proxy forward model and climate model biases on the reconstruction. In the case of proxy models, any biases typically reduce the weight of the proxy in the assimilation, thereby limiting its effect on the reconstruction. We note that improving the accuracy or sophistication of the proxy forward models could increase the influence of many records; for example, transitioning the statistical forward models used here for the PAGES2k sites to more mechanistically accurate proxy system models (Evans et al., 2013) could potentially improve the reconstruction (Dee et al., 2016). However, efforts to develop more complex proxy system models must also exercise caution, as excessive complexity and poorly constrained parameters may lead to overfitting and artificially high skill in the instrumental era at the expense of accuracy during the earlier reconstruction. In this study, we retain the simpler statistical forward models because (1) the PAGES2k proxies are reported to be temperature sensitive (PAGES2k Consortium, 2017), (2) statistical proxy models remain the most common and tractable approach for paleoclimate data assimilation to date (Hakim et al., 2016; Tardif et al., 2019; King et al., 2021), and (3) the simple statistical model minimizes the propagation of climate model biases into the proxy estimates.

With respect to climate models, biases in the mean state can affect proxy estimates

that include parametrizations or thresholds based on absolute units. However, covariance biases are a greater concern, as they introduce errors in the propagation of information from the proxy records to the reconstruction target. In this study, we use a multi-model ensemble (MME) to help reduce the effects of covariance bias from any one model (King et al., 2021; Parsons et al., 2021). However, we note that each model receives equal weight, which effectively treats each model as independent. In reality, many models share common features or code, so this equal weighting may bias an ensemble towards the most similar models (Knutti et al., 2013; Sanderson et al., 2015). For example, the CCSM4 and CESM-LME output used in our MME are both from models developed by the US National Center for Atmospheric Research (NCAR) and may more closely resemble one another than the MPI or MRI models. Future efforts may wish to test different model composition and weights when constructing a MME prior.

Finally, our use of a stationary offline prior implies a stationary estimate of climate system covariance when considered over the full reconstruction period. Although we use a long-term estimate of the SAM's climate covariance, the true covariance may vary on multi-decadal scales (Silvestri and Vera, 2009; Gallant et al., 2013), and these variations will not be captured in our approach. While the assumption of a reasonably stationary covariance is implicitly common to most spatial reconstruction methods (Tingley et al., 2012; Amrhein et al., 2020), the application of transient offline priors (Bhend et al., 2012; Franke et al., 2020; Osman et al., 2021) or online assimilation techniques (Perkins and Hakim, 2017) may enhance future data assimilation reconstruction, although these approaches must balance the utility of evolving covariance estimates with reduced ensemble sizes.

## **B.5 Conclusions**

Our study provides the first reconstruction of the Southern Annular Mode at annual resolution over the entire Common Era. We use a data assimilation method that does not calibrate the proxies directly against the instrumental SAM index, so the reconstruction is not sensitive to observed SAM non-stationarity in the modern era. Our reconstruction

leverages both the SADA and ANZDA in addition to the PAGES2k proxy network and represents a significant increase in paleoclimate information available to reconstruct the SAM. Optimal sensor analysis indicates that the first 1400 years of the reconstruction are strongly influenced by the Oroko and Mt. Read tree-ring chronologies, with additional support from the Plateau Remote, WDC06A, and WDC05A ice cores. As the SADA and ANZDA are added to the proxy network (1400 CE and 1500 CE, respectively), the drought atlases become strong drivers of the reconstruction’s behavior.

Our reconstruction provides a foundation with which to assess the drivers of the SAM’s behavior over the Common Era; such assessments are critical given the SAM’s importance to societies and effects on climate variability throughout the Southern Hemisphere. Although our index and existing SAM reconstructions show limited agreement with one another, all products exhibit the most strongly positive and persistent SAM trend during the last several decades. We find that the modern positive trend in the SAM is outside the range of natural variability over the previous millennium, further confirming a response to anthropogenic forcing. Prior to the most recent decades, we find no relationship between SAM variability and external climate forcing, suggesting that its behavior is dominated by internal variability over the pre-industrial Common Era.

## **B.6 Materials and Methods**

### **B.6.1 Southern Annular Mode Index**

In this study, we use the Gong and Wang (1999) definition of the SAM index:

$$\text{SAM} = P_{40^{\circ}\text{S}}^* - P_{65^{\circ}\text{S}}^* \quad (\text{B.1})$$

where  $P_X^*$  indicates the normalized zonal-mean sea level pressure (SLP) at a particular latitude. The latitudes  $40^{\circ}\text{S}$  and  $65^{\circ}\text{S}$  were selected as the zonal-means with the most strongly anti-correlated SLP anomalies across the mid- and high-latitude Southern Hemisphere. We use this definition, as opposed to an index derived from a principal component analysis because the latitudes of the most strongly anti-correlated SLP anomalies are robust across the climate models considered in our assimilation (Tables B.2, B.3). We

target the austral summer (DJF; December - February) SAM because this corresponds to the seasonality of the climate response of the majority of our proxy network. D18 also suggests that summer SAM reconstructions are more robust to proxy network design than annual reconstructions, which further supports this choice. When calculating the SAM index, we normalize seasonal mean values, rather than individual months. Austral summers span months from two calendar years, and this can introduce date ambiguities for annual records, particularly tree-ring chronologies. Throughout this paper, we use the convention that the year of an austral summer value matches the calendar year of the associated January.

## **B.6.2 Data Products**

### **Reanalysis and Instrumental Indices**

We use monthly precipitation and air-temperature fields from the Twentieth Century Reanalysis V3 (Compo et al., 2011; Slivinski et al., 2019, 20CR) to calibrate our DA method. The 20CR is based on an 80-member ensemble Kalman Filter, and extends from 1850 CE to present at 2 degree resolution. Because of its role in our assimilation method, this effectively sets an upper bound on the resolution of any gridded spatial product used in this reconstruction. We also use the austral summer Marshall Index (Marshall, 2003) and Fogt Index (Fogt et al., 2009; Jones et al., 2009a) to assess the skill of our reconstruction in the modern era. The Marshall Index estimates the Gong and Wang (1999) definition of the SAM (Equation 1), and is based on data from 12 weather stations (6 near 40°S, and 6 near 65°S). Because it uses station data, the Marshall index is not subject to the spurious trends observed in high-latitude Southern Hemisphere reanalysis pressure fields (Marshall, 2003). The Fogt index is constructed using a principal component regression of station pressure data and calibrated to the Marshall index. These indices are commonly used as a comparison point for SAM reconstructions (Villalba et al., 2012; Abram et al., 2014; Dätwyler et al., 2018),

## Climate Proxies

In this reconstruction, we assimilate the PAGES2k temperature-sensitive proxy network (PAGES2k Consortium, 2017), the South American Drought Atlas (SADA; Morales et al., 2020), and the Australia-New Zealand Drought Atlas (ANZDA; Palmer et al., 2015). We limit all three datasets to those sites or locations south of 25°S. Pseudo-proxy tests of other latitude bounds suggests that reconstruction skill is minimally affected by the use of more northward proxy sites and agreement with the instrumental record exhibits a slight maximum for a bound at 25°S (Figure B.7). Overall, this domain maximizes the number of SAM-sensitive proxy sites in our network, while minimizing the effects of distal proxies that primarily reflect other climate signals.

From the PAGES2k dataset, we include all sites from the PAGES2k global temperature reconstruction that have annual or sub-annual temporal resolution. To maintain a common timescale, we bin all sub-annual sites to annual resolution. Our PAGES2k network therefore consists of 40 proxy records: 12 tree-ring chronologies, 3 lake sediment cores, 5 corals, 19 ice-cores, and 1 borehole-derived temperature reconstruction (Table B.4). The tree-ring records are from Tasmania, New Zealand, and the central Andes. The longest two chronologies are from Mt. Read, Tasmania and Oroko, New Zealand, which begin in 494 BCE and 900 CE, respectively; the remaining tree chronologies mostly begin between 1450 CE and 1550 CE. The three lake sediment proxies are derived from the central and southern Andes. The longest record (Laguna Chepical) spans the complete Common Era, while Lagunas Escondida and Aculeo begin in 400 CE and 816 CE, respectively. The five coral records are from the Houtman Abrolhos Islands off the west coast of Australia and begin between 1795 CE and 1900 CE. The Antarctica ice core records have varying temporal coverage. Four sites cover the full Common Era (Plateau Remote, WDC06A, James Ross Island, WAIS-Divide), six more extend at least one millennium, and the remaining nine begin between 1140 CE and 1703 CE. The borehole reconstruction is from WAIS-Divide and begins in 8 CE. For the 40 proxy set, full coverage extends from 1903 CE to 1983 CE with 20 sites remaining by 2000 CE.

The SADA and ANZDA are gridded tree-ring reconstructions of the self-calibrated

Palmer Drought Severity Index (PDSI) during austral summer at annual resolution (Palmer et al., 2015; Morales et al., 2020). The SADA is derived from 286 temperature and precipitation-sensitive tree-ring chronologies and begins in 1400 CE. The atlas covers all of South America south of 12.25°S at 0.5° resolution. Similarly, ANZDA is derived from 176 tree-ring chronologies, as well as one coral record, and begins in 1500 CE. The ANZDA covers Australia east of 136.25°E, and New Zealand, also at 0.5° resolution. The SAM is strongly associated with droughts and pluvials in the domains of both atlases (Morales et al., 2020), supporting their inclusion in our network. Both atlases have significantly higher spatial resolution than the reanalysis data and climate model output used for our reconstruction method. To permit calculations that require the same spatial resolution, we bin both atlases to the lowest resolution spatial grid relevant to a given experiment. For the main reconstruction, after applying latitude screening, our SADA and ANZDA networks consist of 104 and 71 binned records, each on a 2° x 2.5° grid. It is worth noting that several of the PAGES2k tree ring records used in our reconstruction were also used to construct the drought atlases, and these repeat records might initially appear to duplicate information in the reconstruction. However, our Kalman filter method explicitly accounts for covariance between proxy records, and down-weights proxies with repeated information accordingly. Additional details for this process can be found in the following section.

### **B.6.3 Reconstruction Method**

#### **Kalman Filter**

Our reconstruction uses an ensemble Kalman Filter approach (Evensen, 1994, EnKF), which follows the update equation:

$$\mathbf{X}_a = \mathbf{X}_p - \mathbf{K}(\mathbf{Y} - \hat{\mathbf{Y}}) \quad (\text{B.2})$$

in each reconstructed time step. Here, the  $\mathbf{X}_p$  and  $\mathbf{X}_a$  matrices are the initial (prior) and updated (analysis) ensembles of climate model states. Each row holds a target climate variable, and each column a different selection of climate model output (ensemble member).  $\mathbf{Y}$  is a matrix of proxy values for the time step; the columns of  $\mathbf{Y}$  are constant, and

each row holds the value from a particular proxy record repeated once for each ensemble member.  $\hat{\mathbf{Y}}$  holds the model estimates of the proxy values; each row has the estimates for a particular proxy site, and each column has the estimates from a particular ensemble member.  $\mathbf{K}$  is the Kalman gain:

$$\mathbf{K} = \text{cov}(\mathbf{X}_p, \hat{\mathbf{Y}})[\text{cov}(\hat{\mathbf{Y}}) + \mathbf{R}]^{-1} \quad (\text{B.3})$$

where  $\mathbf{R}$  is the matrix of proxy error-covariances. As previously mentioned, the Kalman filter accounts for duplication of information across repeated proxy records. This occurs via the  $\text{cov}(\hat{\mathbf{Y}})$  term in Equation 3, which reduces proxy weights in the Kalman gain as a function of shared proxy covariance. Note that any shared covariance derived from proxies' relationships with the SAM is balanced by the  $\text{cov}(\mathbf{X}_p, \hat{\mathbf{Y}})$  term in Equation B.3. We use a square-root variant of EnKF (Andrews, 1968; Tippett et al., 2003). This modifies equations B.2 and B.3 to update the ensemble mean and deviations separately, and precludes the need for perturbed observations (Whitaker and Hamill, 2002). The Kalman filter can be expressed as a recursive Bayesian filter (Chen et al., 2003; Wikle and Berliner, 2007), so we will often refer to  $\mathbf{X}_p$  and  $\mathbf{X}_a$  as the prior and posterior in this paper.

## Prior

We construct the prior using output from climate models with paleoclimate simulations of the last millennium (Table B.2). We use a multi-model ensemble (MME), which has been found to reduce error relative to single model assimilations (Parsons et al., 2021; King et al., 2021). Our MME consists of CCSM4, CESM-LME, MPI, and MRI, which represent the set of last millennium simulations with spatial resolutions greater than or at the resolution of the 20CR reanalysis. As such, this selection does not require us to bin the drought atlases to lower resolutions than 20CR, which allows us to extract maximum information from SADA and ANZDA. We also tested a larger MME consisting of 10 models with last millennium simulations regardless of resolution. Our tests show that the high-resolution MME maximizes reconstruction skill (Figure B.8). For

CCSM4, MPI, and MRI, we use output from the PMIP3 last1000 (850-1850 CE) and historical (1851-2005 CE) experiments, specifically ensemble member r1i1p1. For CESM-LME, we use output from full-forcing run 2 (850-2005 CE). While the PAGES2k proxy network does include stable oxygen isotope proxies, there are too few high-resolution last millennium isotope-enabled paleoclimate model simulations available to construct a multi-model prior (Parsons et al., 2021).

We use an offline, stationary prior for our assimilation. Offline approaches (Oke et al., 2002; Evensen, 2003) differ from classical Kalman Filters in that updates are not used to inform model simulation. Instead, offline methods use pre-existing model output to build the prior in each time step. The offline approach has been shown to compare favorably with classical (online) methods in paleoclimate contexts but at a fraction of the computational cost (Matsikaris et al., 2015; Acevedo et al., 2017). The stationary prior indicates that we use the same ensemble as the prior for each reconstructed time step. This is common in paleoclimate DA applications (Steiger et al., 2014; Dee et al., 2016; Tardif et al., 2019) and is justified by the limited forecast skill of climate models beyond the annual reconstruction time scale (Bhend et al., 2012). However, stationary priors have been observed to artificially reduce the variability of reconstructions as proxy networks become more sparse (King et al., 2021). Consequently, our use of stationary priors necessitates a correction for the reconstruction’s variability, which is detailed in the methods below.

To build each prior, we first calculate the DJF SAM time-series for each model, normalizing zonal SLP means to the pre-industrial period (850-1849 CE). We then concatenate the SAM index time-series from each model in every year of model output. The final prior has a total of 4624 ensemble members from 4 high-resolution models.

### **Proxy Forward Models and Error Covariances**

The proxy modeling process begins by designing a forward model for each assimilated proxy record. For the PAGES2k records, we follow previous studies (Hakim et al., 2016; King et al., 2021) and use simple univariate linear models:

$$\hat{\mathbf{Y}} = a\mathbf{T} + b \tag{B.4}$$

where  $\hat{\mathbf{Y}}$  is a vector of proxy estimates, and  $\mathbf{T}$  is a vector of seasonal temperature means. Here, the seasonal means used for each site is taken from the seasonal sensitivity reported in the PAGES2k metadata (PAGES2k Consortium, 2017). We determine the coefficients  $a$  and  $b$  by calibrating each proxy PAGES2k record to the corresponding climate data from 20CR. For each proxy site, we first determine the seasonal sensitivity and then linearly regress the proxy record against the seasonal-mean temperature vector from the closest 20CR grid point in all overlapping years from 1950 - 2000 CE. The regression slope and intercept are then used as coefficients  $a$  and  $b$ . For the drought atlases, we estimate proxies by calculating PDSI (Palmer, 1965) using the Thornthwaite estimation of potential evapotranspiration (Thornthwaite, 1948). This uses monthly mean temperature and precipitation from a drought atlas grid cell to compute monthly PDSI values for each year. We then use the austral summer means of these monthly values as the proxy estimates. Effectively:

$$\mathbf{Y} = \text{mean}[ \text{PDSI}_{\text{Thornthwaite}}(\mathbf{T}, \mathbf{P}) ]_{\text{DJF}} \quad (\text{B.5})$$

where  $\mathbf{T}$  and  $\mathbf{P}$  are monthly temperature and precipitation, and  $\mathbf{Y}$  is the drought atlas estimate. We estimate proxy values for the model priors by applying Equations B.4 and B.5 to climate model output and matching each year's estimates to the associated ensemble member in the prior.

Although the PDSI calculation in Equation B.5 uses the Thornthwaite approximation, both drought atlases target an observational dataset based on the Penman-Monteith method (Allen et al., 1994; Palmer et al., 2015; Morales et al., 2020). However, both the Thornthwaite and Penman-Monteith equations have been shown to perform similarly when applied to pre-industrial simulations, and this agreement occurs because the simplifying assumptions of the Thornthwaite method remain valid over the relatively confined range of last millennium temperatures (Smerdon et al., 2015). For the purposes of this study, the Thornthwaite method provides two further advantages: First, the Thornthwaite equation is more computationally tractable, which allows us to apply it to the large spatial regions and the multiple millennium-length climate model simulations used for priors in our assimilation method. Second, because the Thornthwaite calculation requires fewer

climate model data fields to estimate the PDSI (Thornthwaite, 1948; Allen et al., 1994), opportunities for climate model biases to degrade the reconstruction are reduced.

We next estimate the proxy error covariances. These error covariances describe the uncertainty in the comparison of observed records to the proxy estimates ( $\mathbf{Y} - \mathbf{Y}_e$ ). In a classical Kalman Filter, the estimates ( $\hat{\mathbf{Y}}$ ) are known perfectly and this uncertainty is derived from the observations ( $\mathbf{Y}$ ), so  $\mathbf{R}$  is often referred to as observation uncertainty. In paleoclimate contexts, this situation is inverted: proxy measurements are typically precise and uncertainty derives from the simplifications and parameterizations inherent in the estimation equations. Hence, we quantify  $\mathbf{R}$  by running Equations B.4 and B.5 on the 20CR dataset (from 1950-2000 CE) and comparing the estimated proxy values to the real records. The differences between the two sets of values are used to estimate the errors inherent in using simple models and relatively coarse climate data to estimate the temporal behavior of the proxy records. Most EnKF paleoclimate efforts assume that proxy errors are independent, such that  $\mathbf{R}$  is a diagonal matrix (Steiger et al., 2014; Hakim et al., 2016; Steiger et al., 2018; Tardif et al., 2019). This is justified for datasets like PAGES2k, for which proxy uncertainties are dominated by local biological, physical, and mechanistic effects (Jones et al., 2009b). However, the drought atlas grid points are strongly spatially correlated, so this assumption is not appropriate in this study. Instead, we calculate independent error-variances for the proxies in the PAGES2k network, and full error-covariances for both SADA and ANZDA. Hence,  $\mathbf{R}$  is block-diagonal, rather than strictly diagonal. We estimate uncertainty in the final reconstruction from the spread of the assimilation posterior.

### **Variance Correction**

The use of stationary priors creates artifacts in the variability of raw reconstruction. As the proxy network becomes sparse, less information is incorporated in the Kalman Filter, and the updated state is less able to move off the prior mean. This causes reconstruction variability to increase with the size of the proxy network and independently of the climatological record. We apply a variance adjustment scheme to correct for this effect. Variance adjustments are common in paleoclimate reconstructions (Cook et al., 1999; Es-

per et al., 2005; Frank et al., 2007; Anchukaitis et al., 2017) and are inherent to simpler methods like Composite Plus Scale (Mann and Jones, 2003).

Here, we use a series of frozen-network assimilations to adjust temporal variance. There are five sites in our proxy network with observations in every year of the reconstruction. We first assimilate this five-site network over the full interval 1-2000 CE to derive a baseline time-series that is not affected by changes to the proxy network. We next determine each unique set of proxy sites used to update one or more time steps in the reconstruction. We then assimilate each set of proxies over the time steps for which all of the proxies have recorded values, and determine the ratio of this assimilation’s standard deviation to that of the baseline time series over all overlapping years:

$$P(\text{set}) = \sigma_{\text{set}} / \sigma_{\text{Baseline}} \quad (\text{B.6})$$

we then calculate a scaling factor for each time step using the normalized ratio for the associated proxy set:

$$w(t) = P(\text{set}(t)) / \max(P) \quad (\text{B.7})$$

A comparison of the raw and variance-adjusted reconstructions is provided in Figure B.9.

#### **B.6.4 Optimal Sensor Analysis**

We follow a previously established framework for optimal sensor analyses (Comboul et al., 2015). In brief, the method quantifies the ability of proxy sites to reduce the variance of a metric in a posterior ensemble. Here, we use the SAM as our metric, so the optimal sensor analysis here assesses the ability of sites to reduce reconstruction uncertainty in the index. We first compute the total reduction in SAM posterior variance using the complete set of proxies with observations in each time step. We also quantify each site’s ability to reduce reconstruction uncertainty when no other sites are in the proxy network. We refer to this quantity as ‘potential percent constrained variance’.

### B.6.5 External Forcing Analysis

We begin our external forcing analysis by investigating the SAM's response to natural climate forcings. We first use a wavelet coherence analysis to examine the relationship between our SAM reconstruction and a time series of reconstructed solar forcing (Schmidt et al., 2012; Anchukaitis et al., 2017). We next use a superposed epoch analysis (Haurwitz and Brier, 1981) to determine the reconstruction's composite mean response to major volcanic eruptions. We used the eVol2k V3 volcanic forcing dataset (Sigl et al., 2015; Toohey and Sigl, 2017) to select events with a total forcing magnitude greater than or equal to that of Krakatoa. This yielded 28 eruption years: 87, 169, 266, 433, 536, 540, 574, 626, 682, 817, 939, 1108, 1171, 1182, 1230, 1257, 1276, 1286, 1345, 1458, 1600, 1640, 1695, 1783, 1809, 1815, 1831, and 1883. For the SEA, we normalized each event to the mean of the preceding 5 years and examined the composite mean response over the 10 years following volcanic events. We tested the significance of the observed response by bootstrapping 5,000 SEA time series via random draws of 28 event years from the remaining years in the reconstruction.

We next consider the SAM's response to anthropogenic forcings using both our reconstruction and the Marshall index. Before quantifying trends, we first normalize our reconstruction to the Marshall index, such that the mean and variance of the detrended normalized reconstruction matches those of the detrended Marshall index over the years of common overlap (1958-2000 CE). This places the series in the same unit space while preserving differences in the instrumental trend. We then calculate moving trends for the reconstruction over the years 1900-2000 CE using trend window lengths from 31 to 101 years. Similarly, we calculate moving trends for the Marshall Index over the years 1958-2020 CE using trend window lengths from 31 to 63 years. We then use the reconstruction to assess the significance of these trends. For each trend window length, we calculate the distribution of trends with the given window length from the reconstruction over the years 1500-1900 CE. We then use the 90% confidence intervals of these distributions to determine a significance threshold for trends of each length. We also repeat this process using trend distributions from the intervals 1-1900 CE and 1-899 CE to examine the sensitivity

of this analysis to different portions of the reconstruction.

### **Acknowledgements**

We thank David Meko for providing the original PDSI estimation code on which ours is based. This research is supported by a grant from the US National Science Foundation's Paleo Perspectives on Climatic Change program (P2C2) AGS-1803946 to K.J.A. and A.H. T.V. is supported by an Australian Research Council Special Research Initiative for Antarctic Gateway Partnership (SR140300001) and the Australian Antarctic Program Partnership (ASCI000002). K.A. is supported by a grant from the Australian Research Council (FT200100102). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output.

### **Author Contributions**

K.J.A., A.H., and J.K. designed the research. J.K. wrote all the software and code, developed the assimilation method, conducted all calculations, and produced all the figures. A.H., T.V., and K.A. provided guidance and knowledge on individual proxies and Southern Hemisphere climate variability. All authors (J.K., K.J.A., A.H., T.V. and K.A.) interpreted results and wrote the paper.

### **Competing Interests**

The authors declare that they have no conflicts of interest.

### **Data and code availability**

The data and code used to produce this analysis, as well as the final reconstruction, will be made publicly available in a Zenodo repository with a permanent DOI pending review.

## B.7 References

- Abram, N. J., Mulvaney, R., Vimeux, F., Phipps, S. J., Turner, J., and England, M. H.: Evolution of the Southern Annular Mode during the past millennium, *Nature Climate Change*, 4, 564–569, 2014.
- Abram, N. J., Henley, B. J., Sen Gupta, A., Lippmann, T. J., Clarke, H., Dowdy, A. J., Sharples, J. J., Nolan, R. H., Zhang, T., Wooster, M. J., et al.: Connections of climate change and variability to large and extreme forest fires in southeast Australia, *Communications Earth & Environment*, 2, 1–17, 2021.
- Acevedo, W., Fallah, B., Reich, S., and Cubasch, U.: Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model, *Climate of the Past*, 13, 545–557, 2017.
- Allen, R., Smith, M., Perrier, A., and Pereira, L. S.: An update for the definition of reference evapotranspiration, *ICID bulletin*, 43, 1–34, 1994.
- Amrhein, D. E., Hakim, G. J., and Parsons, L. A.: Quantifying structural uncertainty in paleoclimate data assimilation with an application to the last millennium, *Geophysical Research Letters*, 47, e2020GL090485, 2020.
- Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D’Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B. E., et al.: Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions, *Quaternary Science Reviews*, 163, 1–22, 2017.
- Andrews, A.: A square root formulation of the Kalman covariance equations., *AIAA Journal*, 6, 1165–1166, 1968.
- Arblaster, J. M. and Meehl, G. A.: Contributions of external forcings to southern annular mode trends, *Journal of Climate*, 19, 2896–2905, 2006.

- Barrucand, M. G., Zitto, M. E., Piotrkowski, R., Canziani, P., and O'Neill, A.: Historical SAM index time series: linear and nonlinear analysis, *International Journal of Climatology*, 38, e1091–e1106, 2018.
- Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, *Climate of the Past*, 8, 963–976, 2012.
- Bradley, R. S.: Are there optimum sites for global paleotemperature reconstruction?, in: *Climatic variations and forcing mechanisms of the last 2000 years*, pp. 603–624, Springer, 1996.
- Brönnimann, S., Jacques-Coper, M., Rozanov, E., Fischer, A. M., Morgenstern, O., Zeng, G., Akiyoshi, H., and Yamashita, Y.: Tropical circulation and precipitation response to ozone depletion and recovery, *Environmental Research Letters*, 12, 064 011, 2017.
- Byrne, N. J., Shepherd, T. G., and Polichtchouk, I.: Subseasonal-to-seasonal predictability of the Southern Hemisphere eddy-driven jet during austral spring and early summer, *Journal of Geophysical Research: Atmospheres*, 124, 6841–6855, 2019.
- Cai, W. and Cowan, T.: SAM and regional rainfall in IPCC AR4 models: Can anthropogenic forcing account for southwest Western Australian winter rainfall reduction?, *Geophysical Research Letters*, 33, 2006.
- Cai, W., Van Rensch, P., Borlace, S., and Cowan, T.: Does the Southern Annular Mode contribute to the persistence of the multidecade-long drought over southwest Western Australia?, *Geophysical Research Letters*, 38, 2011.
- Chen, Z. et al.: Bayesian filtering: From Kalman filters to particle filters, and beyond, *Statistics*, 182, 1–69, 2003.
- Comboul, M., Emile-Geay, J., Hakim, G. J., and Evans, M. N.: Paleoclimate sampling as a sensor placement problem, *Journal of Climate*, 28, 7717–7740, 2015.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al.: The twentieth century

- reanalysis project, *Quarterly Journal of the Royal Meteorological Society*, 137, 1–28, 2011.
- Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K.: Drought reconstructions for the continental United States, *Journal of Climate*, 12, 1145–1162, 1999.
- Dalaiden, Q., Goosse, H., Rezsöhazy, J., and Thomas, E. R.: Reconstructing atmospheric circulation and sea-ice extent in the West Antarctic over the past 200 years using data assimilation, *Climate Dynamics*, pp. 1–25, 2021.
- Dätwyler, C., Neukom, R., Abram, N. J., Gallant, A. J., Grosjean, M., Jacques-Coper, M., Karoly, D. J., and Villalba, R.: Teleconnection stationarity, variability and trends of the Southern Annular Mode (SAM) during the last millennium, *Climate Dynamics*, 51, 2321–2339, 2018.
- Dee, S. G., Steiger, N. J., Emile-Geay, J., and Hakim, G. J.: On the utility of proxy system models for estimating climate states over the Common Era, *Journal of Advances in Modeling Earth Systems*, 8, 1164–1179, 2016.
- Ding, Q., Steig, E. J., Battisti, D. S., and Wallace, J. M.: Influence of the tropics on the southern annular mode, *Journal of Climate*, 25, 6330–6348, 2012.
- Emile-Geay, J., Cobb, K. M., Mann, M. E., and Wittenberg, A. T.: Estimating central equatorial Pacific SST variability over the past millennium. Part II: Reconstructions and implications, *Journal of Climate*, 26, 2329–2352, 2013.
- England, M. R., Polvani, L. M., Smith, K. L., Landrum, L., and Holland, M. M.: Robust response of the Amundsen Sea Low to stratospheric ozone depletion, *Geophysical Research Letters*, 43, 8207–8213, 2016.
- Esper, J., Frank, D. C., Wilson, R. J., and Briffa, K. R.: Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophysical Research Letters*, 32, 2005.

- Evans, M. N., Kaplan, A., and Cane, M. A.: Optimal sites for coral-based reconstruction of global sea surface temperature, *Paleoceanography*, 13, 502–516, 1998.
- Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Science Reviews*, 76, 16–28, 2013.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, 99, 10 143–10 162, 1994.
- Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343–367, 2003.
- Fogt, R. L. and Bromwich, D. H.: Decadal variability of the ENSO teleconnection to the high-latitude South Pacific governed by coupling with the southern annular mode, *Journal of Climate*, 19, 979–997, 2006.
- Fogt, R. L. and Marshall, G. J.: The Southern Annular Mode: variability, trends, and climate impacts across the Southern Hemisphere, *Wiley Interdisciplinary Reviews: Climate Change*, 11, e652, 2020.
- Fogt, R. L., Perlwitz, J., Monaghan, A. J., Bromwich, D. H., Jones, J. M., and Marshall, G. J.: Historical SAM variability. Part II: Twentieth-century variability and trends from reconstructions, observations, and the IPCC AR4 models, *Journal of Climate*, 22, 5346–5365, 2009.
- Fogt, R. L., Bromwich, D. H., and Hines, K. M.: Understanding the SAM influence on the South Pacific ENSO teleconnection, *Climate Dynamics*, 36, 1555–1576, 2011.
- Fogt, R. L., Goergens, C. A., Jones, J. M., Schneider, D. P., Nicolas, J. P., Bromwich, D. H., and Dusselier, H. E.: A twentieth century perspective on summer Antarctic pressure change and variability and contributions from tropical SSTs and ozone depletion, *Geophysical Research Letters*, 44, 9918–9927, 2017.

- Frank, D., Esper, J., and Cook, E. R.: Adjustment for proxy number and coherence in a large-scale temperature reconstruction, *Geophysical Research Letters*, 34, 2007.
- Franke, J., Valler, V., Brönnimann, S., Neukom, R., and Jaume-Santero, F.: The importance of input data quality and quantity in climate field reconstructions—results from the assimilation of various tree-ring collections, *Climate of the Past*, 16, 1061–1074, 2020.
- Gallant, A. J., Phipps, S. J., Karoly, D. J., Mullan, A. B., and Lorrey, A. M.: Nonstationary Australasian teleconnections and implications for paleoclimate reconstructions, *Journal of Climate*, 26, 8827–8849, 2013.
- Gillett, N. P. and Thompson, D. W.: Simulation of recent Southern Hemisphere climate change, *Science*, 302, 273–275, 2003.
- Gillett, N. P., Kell, T. D., and Jones, P.: Regional climate impacts of the Southern Annular Mode, *Geophysical Research Letters*, 33, 2006.
- Gillett, N. P., Fyfe, J. C., and Parker, D. E.: Attribution of observed sea level pressure trends to greenhouse gas, aerosol, and ozone changes, *Geophysical Research Letters*, 40, 2302–2306, 2013.
- Gong, D. and Wang, S.: Definition of Antarctic oscillation index, *Geophysical Research Letters*, 26, 459–462, 1999.
- Goosse, H., Guiot, J., Mann, M. E., Dubinkina, S., and Sallaz-Damaz, Y.: The Medieval Climate Anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation, *Global and Planetary Change*, 84, 35–47, 2012.
- Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., Gong, D., Kaufman, D. S., Nnamchi, H. C., Quaas, J., Rivera, J. A., Sathyendranath, S., Smith, S. L., Trewin, B., von Shuckmann, K., and Vose, R. S.: Changing State of the Climate System, in: *Climate Change 2021: The Physical Science Basis. Contribution*

- of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., chap. 2, Cambridge University Press, 2021.
- Gupta, A. S. and England, M. H.: Coupled ocean–atmosphere–ice response to variations in the southern annular mode, *Journal of Climate*, 19, 4457–4486, 2006.
- Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The Last Millennium Climate Reanalysis project: Framework and first results, *Journal of Geophysical Research: Atmospheres*, 121, 6745–6764, 2016.
- Hall, A. and Visbeck, M.: Synchronous variability in the Southern Hemisphere atmosphere, sea ice, and ocean resulting from the annular mode, *Journal of Climate*, 15, 3043–3057, 2002.
- Haurwitz, M. W. and Brier, G. W.: A critique of the superposed epoch analysis method: its application to solar–weather relations, *Monthly Weather Review*, 109, 2074–2079, 1981.
- Hendon, H. H., Thompson, D. W., and Wheeler, M. C.: Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode, *Journal of Climate*, 20, 2452–2467, 2007.
- Hessl, A., Allen, K. J., Vance, T., Abram, N. J., and Saunders, K. M.: Reconstructions of the southern annular mode (SAM) during the last millennium, *Progress in Physical Geography*, 41, 834–849, 2017.
- Ho, M., Kiem, A., and Verdon-Kidd, D.: The Southern Annular Mode: a comparison of indices, *Hydrology and Earth System Sciences*, 16, 967–982, 2012.
- Holz, A. and Veblen, T. T.: Variability in the Southern Annular Mode determines wildfire activity in Patagonia, *Geophysical Research Letters*, 38, 2011.

- Holz, A., Paritsis, J., Mundo, I. A., Veblen, T. T., Kitzberger, T., Williamson, G. J., Araóz, E., Bustos-Schindler, C., González, M. E., Grau, H. R., et al.: Southern Annular Mode drives multicentury wildfire activity in southern South America, *Proceedings of the National Academy of Sciences*, 114, 9552–9557, 2017.
- Huiskamp, W. and McGregor, S.: Quantifying Southern Annular Mode paleoreconstruction skill in a model framework, *Climate of the Past*, 17, 1819–1839, 2021.
- Jones, J. M., Fogt, R. L., Widmann, M., Marshall, G. J., Jones, P. D., and Visbeck, M.: Historical SAM variability. Part I: Century-length seasonal reconstructions, *Journal of Climate*, 22, 5319–5345, 2009a.
- Jones, J. M., Gille, S. T., Goosse, H., Abram, N. J., Canziani, P. O., Charman, D. J., Clem, K. R., Crosta, X., de Lavergne, C., Eisenman, I., et al.: Assessing recent trends in high-latitude Southern Hemisphere surface climate, *Nature Climate Change*, 6, 917–926, 2016.
- Jones, P. D., Briffa, K. R., Osborn, T., Lough, J. M., van Ommen, T. D., Vinther, B. M., Luterbacher, J., Wahl, E., Zwiers, F., Mann, M. E., et al.: High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects, *The Holocene*, 19, 3–49, 2009b.
- Kidston, J., Renwick, J., and McGregor, J.: Hemispheric-scale seasonality of the Southern Annular Mode and impacts on the climate of New Zealand, *Journal of Climate*, 22, 4759–4770, 2009.
- King, J. M., Anchukaitis, K. J., Tierney, J. E., Hakim, G. J., Emile-Geay, J., Zhu, F., and Wilson, R.: A data assimilation approach to last millennium temperature field reconstruction using a limited high-sensitivity proxy network, *Journal of Climate*, pp. 1–64, 2021.
- Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, *Geophysical Research Letters*, 40, 1194–1199, 2013.

- Kohyama, T. and Hartmann, D. L.: Antarctic sea ice response to weather and climate modes of variability, *Journal of Climate*, 29, 721–741, 2016.
- Kwok, R. and Comiso, J. C.: Spatial patterns of variability in Antarctic surface temperature: Connections to the Southern Hemisphere Annular Mode and the Southern Oscillation, *Geophysical Research Letters*, 29, 50–1, 2002.
- L’Heureux, M. L. and Thompson, D. W.: Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation, *Journal of Climate*, 19, 276–287, 2006.
- Lim, E.-P., Hendon, H., and Thompson, D.: Seasonal evolution of stratosphere-troposphere coupling in the Southern Hemisphere and implications for the predictability of surface climate, *Journal of Geophysical Research: Atmospheres*, 123, 12–002, 2018.
- Lim, E.-P., Hendon, H. H., Butler, A. H., Thompson, D. W., Lawrence, Z. D., Scaife, A. A., Shepherd, T. G., Polichtchouk, I., Nakamura, H., Kobayashi, C., et al.: The 2019 Southern Hemisphere stratospheric polar vortex weakening and its impacts, *Bulletin of the American Meteorological Society*, 102, E1150–E1171, 2021.
- Lovenduski, N. S., Gruber, N., Doney, S. C., and Lima, I. D.: Enhanced CO<sub>2</sub> outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, *Global Biogeochemical Cycles*, 21, 2007.
- Lovenduski, N. S., Fay, A. R., and McKinley, G. A.: Observing multidecadal trends in Southern Ocean CO<sub>2</sub> uptake: What can we learn from an ocean model?, *Global Biogeochemical Cycles*, 29, 416–426, 2015.
- Mann, M. E. and Jones, P. D.: Global surface temperatures over the past two millennia, *Geophysical Research Letters*, 30, 2003.
- Mariani, M. and Fletcher, M.-S.: The Southern Annular Mode determines interannual and centennial-scale fire activity in temperate southwest Tasmania, Australia, *Geophysical Research Letters*, 43, 1702–1709, 2016.

- Marshall, G. J.: Trends in the Southern Annular Mode from observations and reanalyses, *Journal of Climate*, 16, 4134–4143, 2003.
- Marshall, G. J., Orr, A., Van Lipzig, N. P., and King, J. C.: The impact of a changing Southern Hemisphere Annular Mode on Antarctic Peninsula summer temperatures, *Journal of Climate*, 19, 5388–5404, 2006.
- Matsikaris, A., Widmann, M., and Jungclaus, J. H.: On-line and off-line data assimilation in palaeoclimatology: a case study, *Climate of the Past*, 11, 81–93, 2015.
- Mauger, G., Bumbaco, K., Hakim, G., and Mote, P.: Optimal design of a climatological network: Beyond practical considerations, *Geoscientific Instrumentation, Methods and Data Systems*, 2, 199–212, 2013.
- Morales, M. S., Cook, E. R., Barichivich, J., Christie, D. A., Villalba, R., LeQuesne, C., Srur, A. M., Ferrero, M. E., González-Reyes, Á., Couvreur, F., et al.: Six hundred years of South American tree rings reveal an increase in severe hydroclimatic events since mid-20th century, *Proceedings of the National Academy of Sciences*, 117, 16 816–16 823, 2020.
- Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P.: No evidence for globally coherent warm and cold periods over the preindustrial Common Era, *Nature*, 571, 550–554, 2019.
- Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., and Kosro, P. M.: Assimilation of surface velocity data into a primitive equation coastal ocean model, *Journal of Geophysical Research: Oceans*, 107, <https://doi.org/10.1029/2000JC000511>, 2002.
- Osman, M. B., Tierney, J. E., Zhu, J., Tardif, R., Hakim, G. J., King, J., and Poulsen, C. J.: Globally resolved surface temperatures since the Last Glacial Maximum, *Nature*, 599, 239–244, 2021.
- PAGES 2k-PMIP3 group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, *Climate of the Past*, 11, 1673–1699, <https://doi.org/10.5194/cp-11-1673-2015>, 2015.

- PAGES2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, *Scientific Data*, 4, 2017.
- Palmer, J. G., Cook, E. R., Turney, C. S., Allen, K., Fenwick, P., Cook, B. I., O'Donnell, A., Lough, J., Grierson, P., and Baker, P.: Drought variability in the eastern Australia and New Zealand summer drought atlas (ANZDA, CE 1500–2012) modulated by the Interdecadal Pacific Oscillation, *Environmental Research Letters*, 10, 124002, 2015.
- Palmer, W. C.: *Meteorological drought*, vol. 30, US Department of Commerce, Weather Bureau, 1965.
- Parsons, L. A., Amrhein, D. E., Sanchez, S. C., Tardif, R., Brennan, M. K., and Hakim, G. J.: Do Multi-Model Ensembles Improve Reconstruction Skill in Paleoclimate Data Assimilation?, *Earth and Space Science*, 8, e2020EA001467, 2021.
- Perkins, W. A. and Hakim, G. J.: Reconstructing paleoclimate fields using online data assimilation with a linear inverse model, *Climate of the Past*, 13, 421–436, 2017.
- Polvani, L. M., Waugh, D. W., Correa, G. J., and Son, S.-W.: Stratospheric ozone depletion: The main driver of twentieth-century atmospheric circulation changes in the Southern Hemisphere, *Journal of Climate*, 24, 795–812, 2011.
- Reason, C. and Rouault, M.: Links between the Antarctic Oscillation and winter rainfall over western South Africa, *Geophysical Research Letters*, 32, 2005.
- Rogers, J. C. and Van Loon, H.: Spatial variability of sea level pressure and 500 mb height anomalies over the Southern Hemisphere, *Monthly Weather Review*, 110, 1375–1392, 1982.
- Sallée, J.-B., Speer, K., and Rintoul, S.: Zonally asymmetric response of the Southern Ocean mixed-layer depth to the Southern Annular Mode, *Nature Geoscience*, 3, 273–279, 2010.
- Sanderson, B. M., Knutti, R., and Caldwell, P.: A representative democracy to reduce interdependency in a multimodel ensemble, *Journal of Climate*, 28, 5171–5194, 2015.

- Schmidt, G., Jungclauss, J. H., Ammann, C., Bard, E., Braconnot, P., Crowley, T., Delaygue, G., Joos, F., Krivova, N., Muscheler, R., et al.: Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1. 1), *Geoscientific Model Development*, 5, 185–191, 2012.
- Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., et al.: Timing and climate forcing of volcanic eruptions for the past 2,500 years, *Nature*, 523, 543–549, 2015.
- Silvestri, G. and Vera, C.: Nonstationary impacts of the southern annular mode on Southern Hemisphere climate, *Journal of Climate*, 22, 6142–6148, 2009.
- Silvestri, G. E. and Vera, C. S.: Antarctic Oscillation signal on precipitation anomalies over southeastern South America, *Geophysical Research Letters*, 30, 2003.
- Simpkins, G. R., Ciasto, L. M., Thompson, D. W., and England, M. H.: Seasonal relationships between large-scale climate variability and Antarctic sea ice concentration, *Journal of Climate*, 25, 5451–5469, 2012.
- Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., et al.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, *Quarterly Journal of the Royal Meteorological Society*, 145, 2876–2908, 2019.
- Smerdon, J. E., Cook, B. I., Cook, E. R., and Seager, R.: Bridging past and future climate across paleoclimatic reconstructions, observations, and models: A hydroclimate case study, *Journal of Climate*, 28, 3212–3231, 2015.
- Sousa, P. M., Blamey, R. C., Reason, C. J., Ramos, A. M., and Trigo, R. M.: The 'Day Zero' Cape Town drought and the poleward migration of moisture corridors, *Environmental Research Letters*, 13, 124025, 2018.
- Stammerjohn, S. E., Martinson, D., Smith, R., Yuan, X., and Rind, D.: Trends in Antarctic annual sea ice retreat and advance and their relation to El Niño–Southern Oscillation

- and Southern Annular Mode variability, *Journal of Geophysical Research: Oceans*, 113, 2008.
- Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of time-averaged pseudoproxies for climate reconstruction, *Journal of Climate*, 27, 426–441, 2014.
- Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, *Scientific Data*, 5, <https://doi.org/10.1086/sdata.2018.86>, 2018.
- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling., *Climate of the Past*, 15, 1251–1273, 2019.
- Thompson, D. W. and Solomon, S.: Interpretation of recent Southern Hemisphere climate change, *Science*, 296, 895–899, 2002.
- Thompson, D. W. and Wallace, J. M.: Annular modes in the extratropical circulation. Part I: Month-to-month variability, *Journal of Climate*, 13, 1000–1016, 2000.
- Thompson, D. W., Wallace, J. M., and Hegerl, G. C.: Annular modes in the extratropical circulation. Part II: Trends, *Journal of Climate*, 13, 1018–1036, 2000.
- Thompson, D. W., Solomon, S., Kushner, P. J., England, M. H., Grise, K. M., and Karoly, D. J.: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change, *Nature Geoscience*, 4, 741–749, 2011.
- Thornthwaite, C. W.: An approach toward a rational classification of climate, *Geographical Review*, 38, 55–94, 1948.
- Tierney, J. E., Abram, N. J., Anchukaitis, K. J., Evans, M. N., Giry, C., Kilbourne, K. H., Saenger, C. P., Wu, H. C., and Zinke, J.: Tropical sea surface temperatures for the

- past four centuries reconstructed from coral archives, *Paleoceanography*, 30, 226–252, 2015.
- Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., and Poulsen, C. J.: Glacial cooling and climate sensitivity revisited, *Nature*, 584, 569–573, 2020.
- Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, *Quaternary Science Reviews*, 35, 1–22, 2012.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble square root filters, *Monthly Weather Review*, 131, 1485–1490, 2003.
- Toohey, M. and Sigl, M.: Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE, *Earth System Science Data*, 9, 809–831, 2017.
- Van Lipzig, N. P., Marshall, G. J., Orr, A., and King, J. C.: The relationship between the Southern Hemisphere Annular Mode and Antarctic Peninsula summer temperatures: Analysis of a high-resolution model climatology, *Journal of Climate*, 21, 1649–1668, 2008.
- Verdon-Kidd, D. C. and Kiem, A. S.: Nature and causes of protracted droughts in south-east Australia: Comparison between the Federation, WWII, and Big Dry droughts, *Geophysical Research Letters*, 36, 2009.
- Villalba, R., Lara, A., Masiokas, M. H., Urrutia, R., Luckman, B. H., Marshall, G. J., Mundo, I. A., Christie, D. A., Cook, E. R., Neukom, R., et al.: Unusual Southern Hemisphere tree growth patterns induced by changes in the Southern Annular Mode, *Nature Geoscience*, 5, 793–798, 2012.
- Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, 130, 1913–1924, 2002.
- Wikle, C. K. and Berliner, L. M.: A Bayesian tutorial for data assimilation, *Physica D: Nonlinear Phenomena*, 230, 1–16, 2007.

- Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., et al.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, *Quaternary Science Reviews*, 134, 1–18, 2016.
- Wright, N. M., Krause, C. E., Phipps, S. J., Boschat, G., and Abram, N. J.: Influence of long-term changes in solar irradiance forcing on the Southern Annular Mode, *Climate of the Past Discussions*, pp. 1–32, 2021.
- Yun, K.-S. and Timmermann, A.: Decadal monsoon-ENSO relationships reexamined, *Geophysical Research Letters*, 45, 2014–2021, 2018.
- Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., and Anchukaitis, K. J.: Resolving the differences in the simulated and reconstructed temperature response to volcanism, *Geophysical Research Letters*, 47, e2019GL086908, 2020.
- Zhu, F., Emile-Geay, J., Anchukaitis, K. J., Hakim, G. J., Wittenberg, A. T., Morales, M. S., Toohey, M., and King, J.: A re-appraisal of the ENSO response to volcanism with paleoclimate data assimilation, *Nature Communications*, 13, <https://doi.org/10.1038/s41467-022-28210-1>, 2022.

Metric	Marshall Index (1958-2000)	Fogt Index (1958-2000)	Fogt Index (1866-2000)
Correlation ( $p \ll 0.001$ )	0.72	0.67	0.56
RMSE	1.45	1.56	1.80
$\sigma$ Ratio	0.97	1.03	1.15
Mean Bias	-0.26	0.45	-0.29

Table B.1: Reconstruction Skill

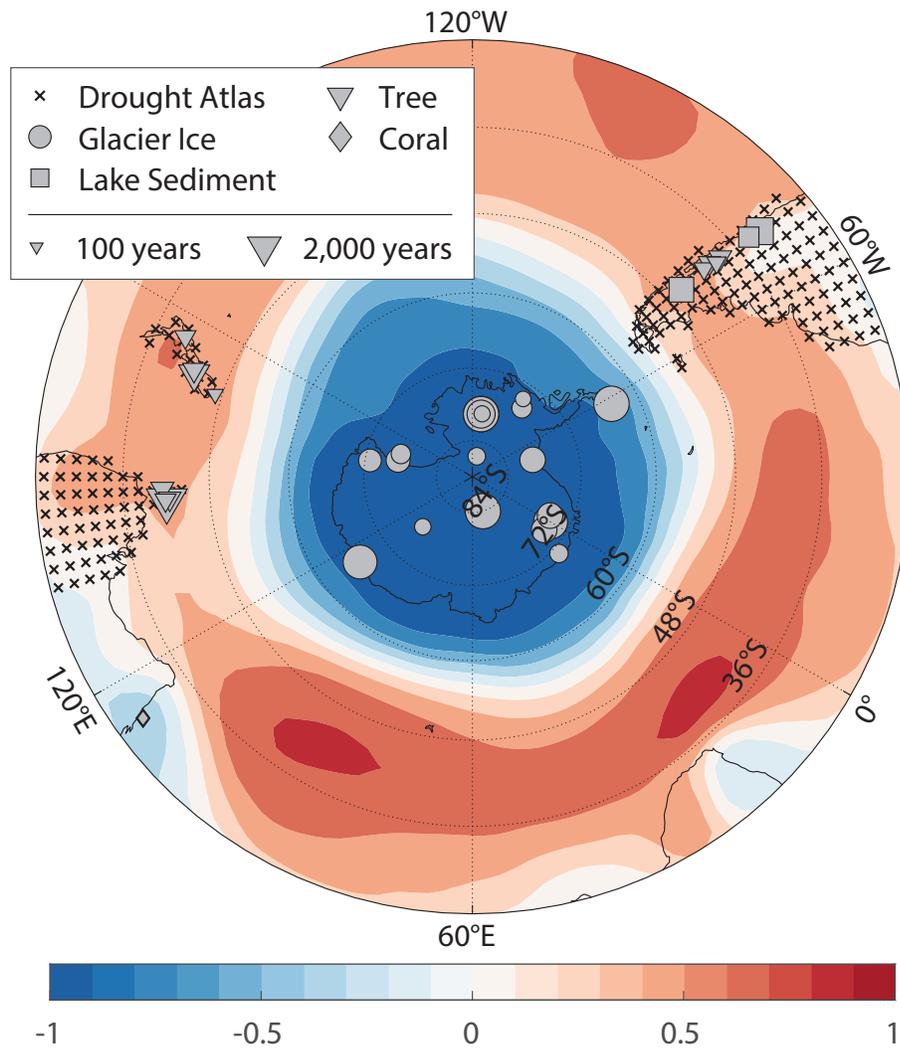


Figure B.1: Map of the proxy network. Black Xs indicate the centroid of binned drought atlas sites. Grey markers indicate PAGES2k sites. The size of the PAGES2k markers correspond to the length of each record. Filled color contours show the field correlation between the 20CR SAM index with 20CR DJF sea level pressure over the period 1958-2000 CE.

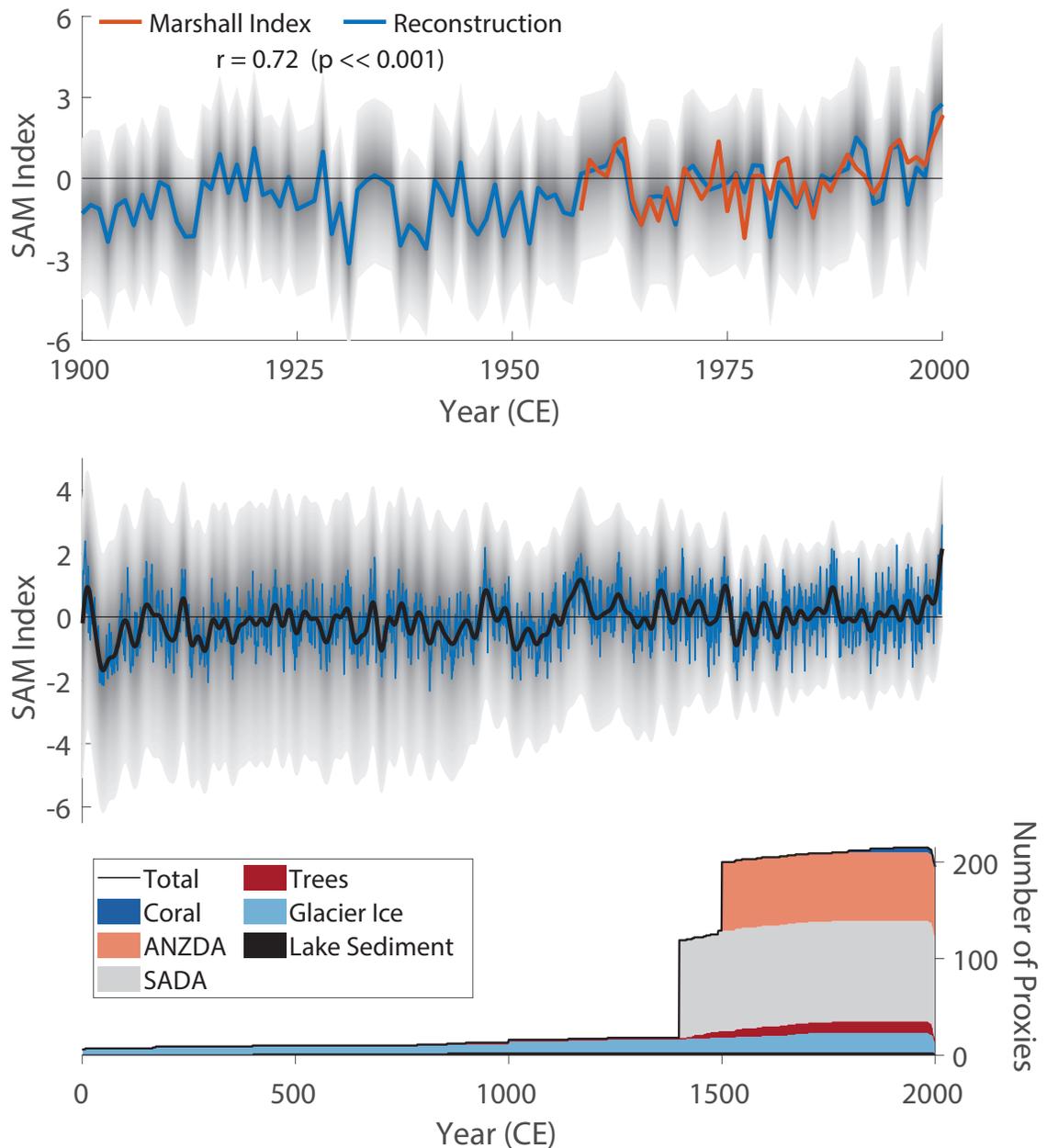


Figure B.2: Evolution of the reconstruction over time. Top: Comparison of the annual reconstruction (blue) with the Marshall index (red) over the instrumental era. Shading indicates the 5-95 percentiles of the reconstruction. Middle: Evolution of the annual reconstruction (blue) and 31-year lowpass filtered (black) over the Common Era. Shading indicates the 5-95 percentiles of the lowpass filtered series. Bottom: Composition of the proxy network over time. Colors for proxy types are as follows: Dark blue (coral), pink (ANZDA), grey (SADA), dark red (trees), light blue (glacier ice), black (lake sediment).

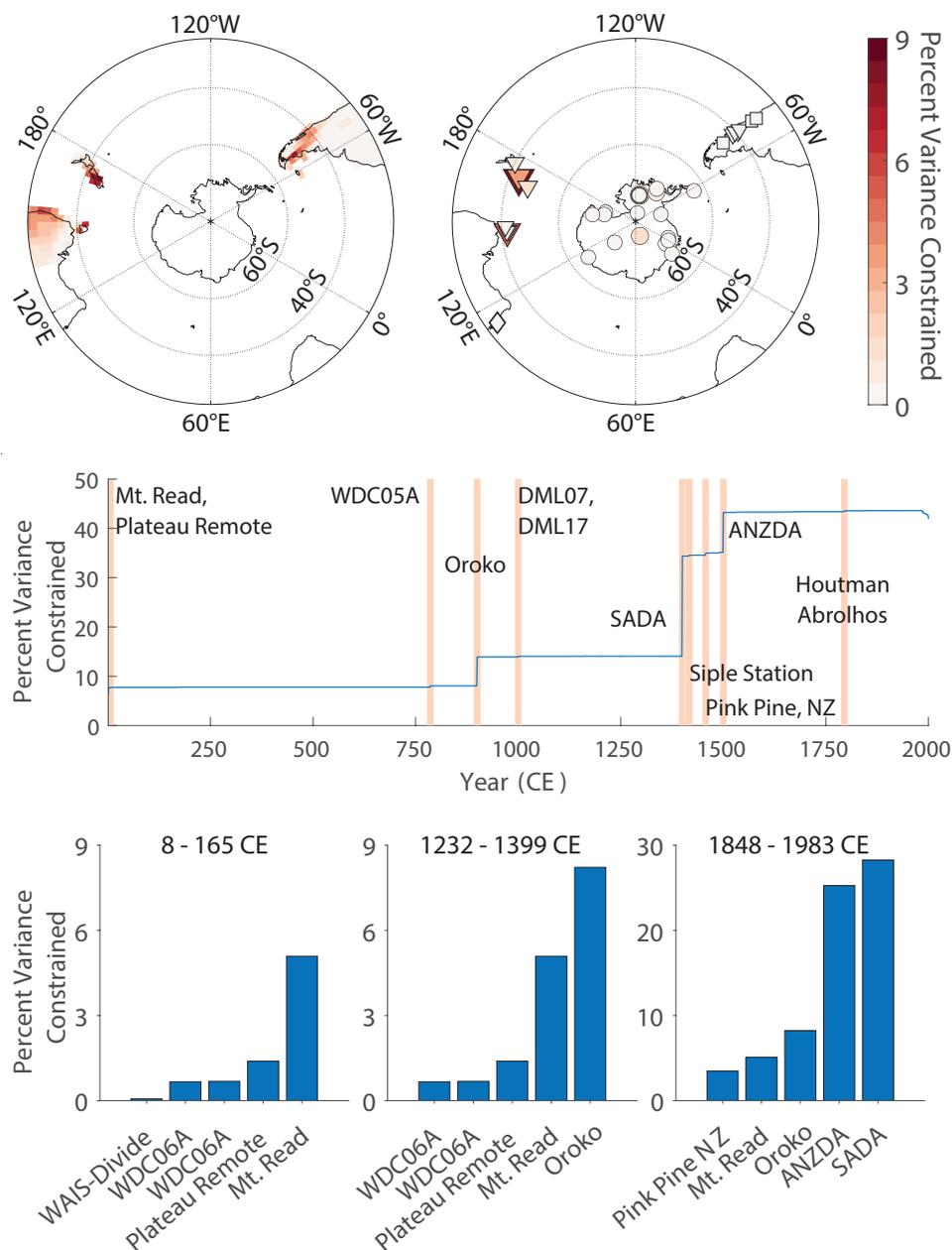


Figure B.3: Optimal Sensor Analysis. Top: Maps of the potential ability for drought atlas (left) and PAGES2k (right) sites to constrain reconstruction posterior variance. Middle: Evolution of reconstruction posterior variance over time. Yellow bars indicate the addition of the indicated proxy to the network. Bottom: Ranked histograms of the seven sites with greatest potential influence in the early reconstruction (left), immediately before the addition of drought atlases (middle) and for the full network (right). Potential influence is determined as the uncertainty constrained by a single-proxy network.

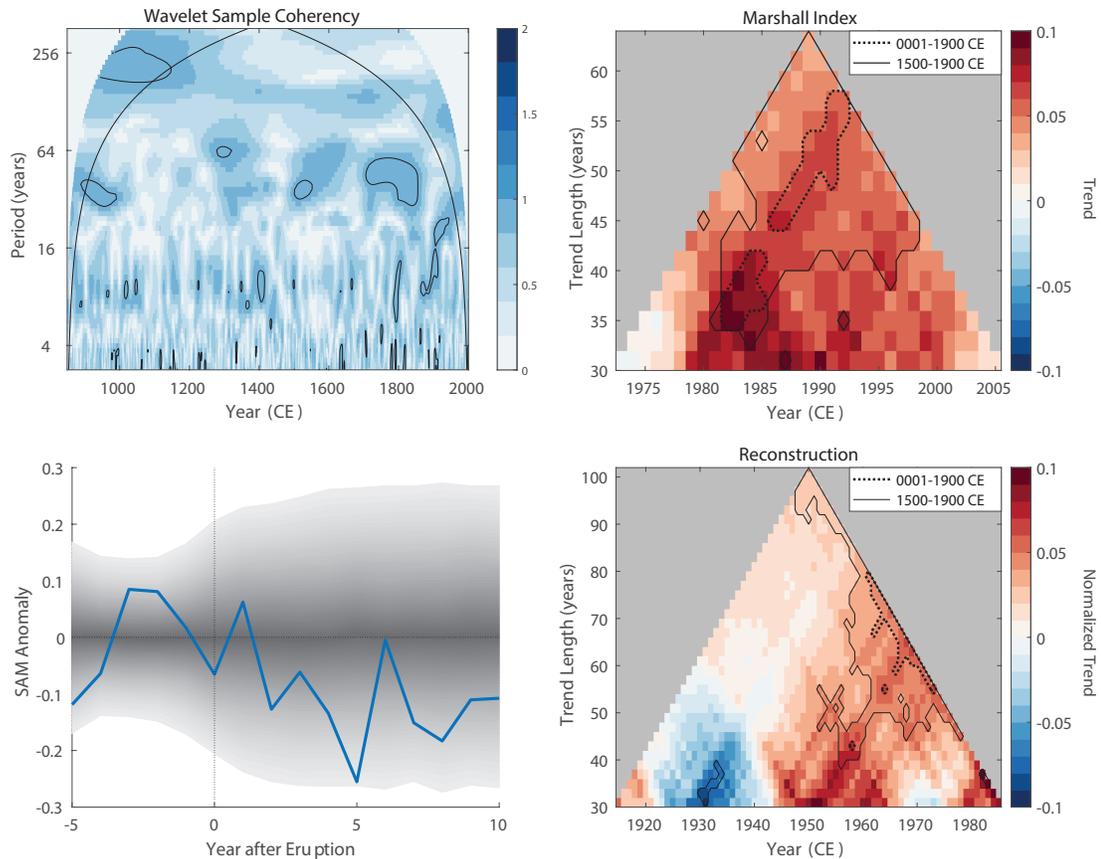


Figure B.4: SAM climate responses. Top left: Wavelet coherence of the reconstructed SAM index with the solar forcing reconstruction (Schmidt et al., 2012). Bottom left: Composite mean response to major volcanic eruptions. Shading indicates 5-95 percentiles. Blue line is the ensemble mean. Right: Instrumental SAM trends for the Marshall Index (top right) and reconstruction (bottom right). Colored points indicate trends calculated from a sliding window centered on the given year. Solid (dotted) contours surround statistically significant trends at the 90% confidence interval relative to the reconstruction over the period 1500-1900 CE (1-1900 CE).

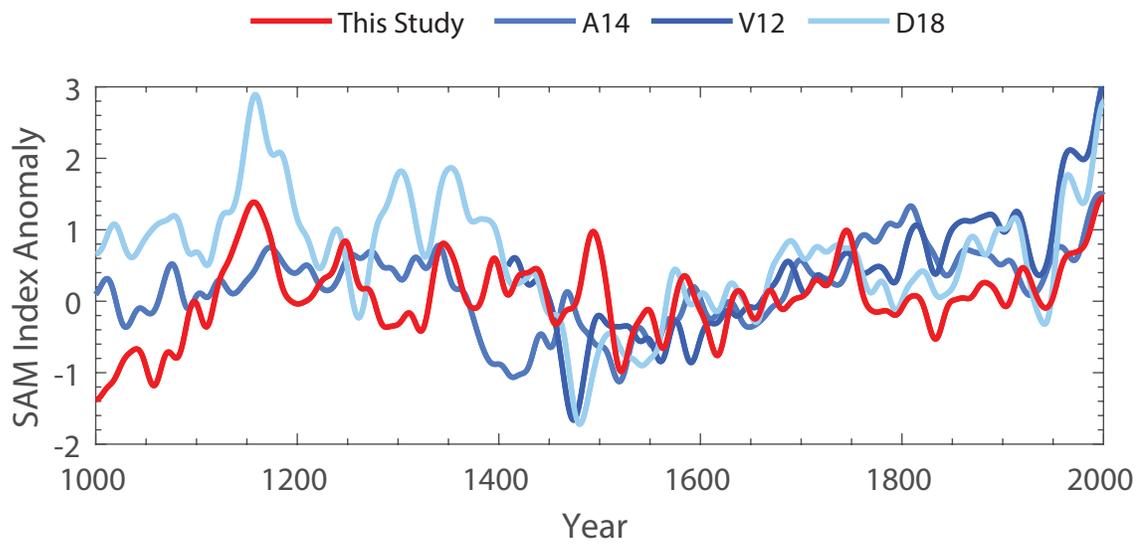


Figure B.5: Comparison of SAM reconstructions over the last millennium. All reconstructions are smoothed via a 30-year Gaussian filter and normalized to the period 1400-1850 CE.

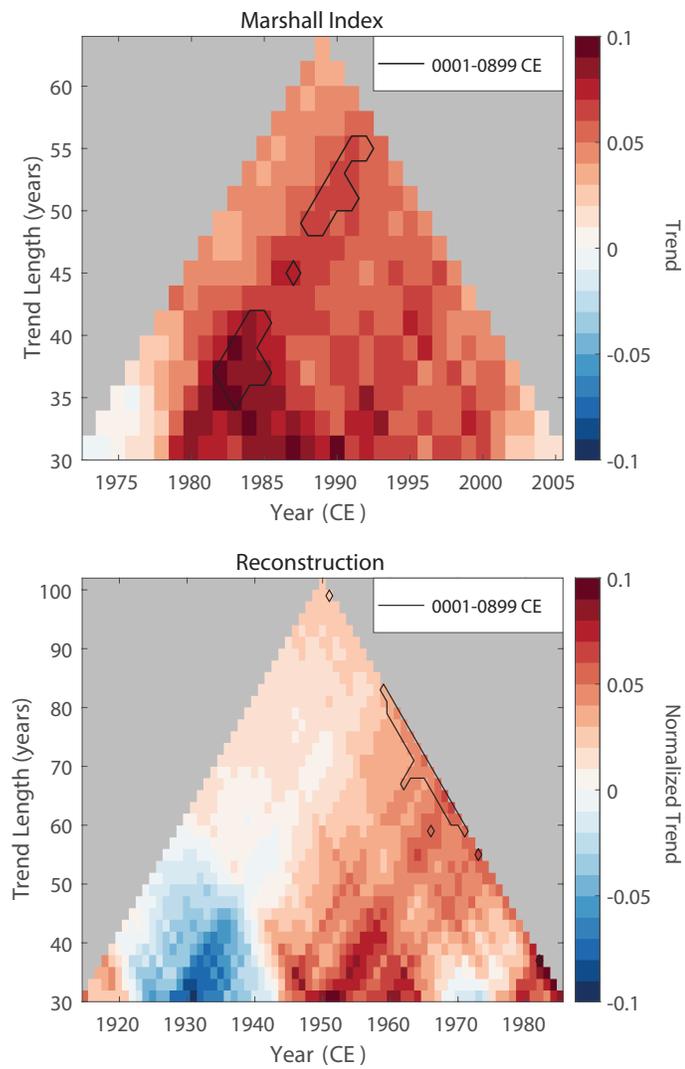


Figure B.6: Trend analysis as in Figure B.4, but using the early reconstruction (1-899 CE) to quantify natural variability and assess trend significance.

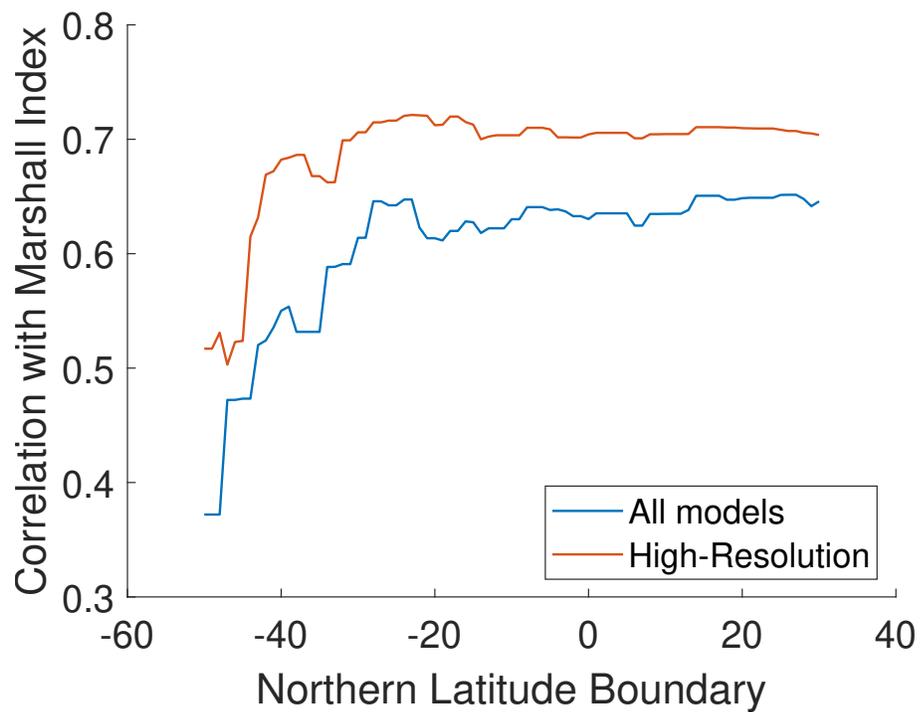


Figure B.7: Correlation of reconstructions with the Marshall Index for different latitude cutoff boundaries. Correlation is assessed over the period 1958-2000 CE. Blue line shows results for reconstructions with an MME prior constructed from all 10 models. Red line shows results for reconstructions with an MME prior built from the 4 high-resolution models.

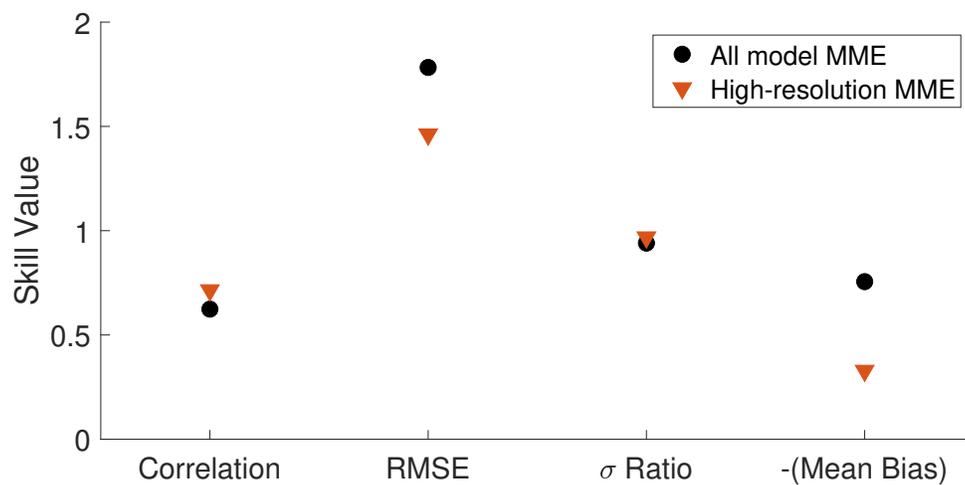


Figure B.8: Skill comparison for different priors. Skill metrics are assessed relative to the Marshall Index over the period 1958-2000. Prior to skill assessment, reconstruction time series are scaled such that the mean and standard deviation of the detrended reconstructions match the mean and standard deviation of the detrended Marshall Index. Black circles indicate values for the all-model MME prior (10 models). Red triangles indicate values for the high-resolution MME prior (4 models). Figure columns are 1. Correlation, 2. Root mean-square error, 3. Standard deviation ratios (computed as the ratio of reconstruction variability over Marshall Index variability, and 4. Bias in the mean value of the series. To condense the plot, negative mean biases are shown here.

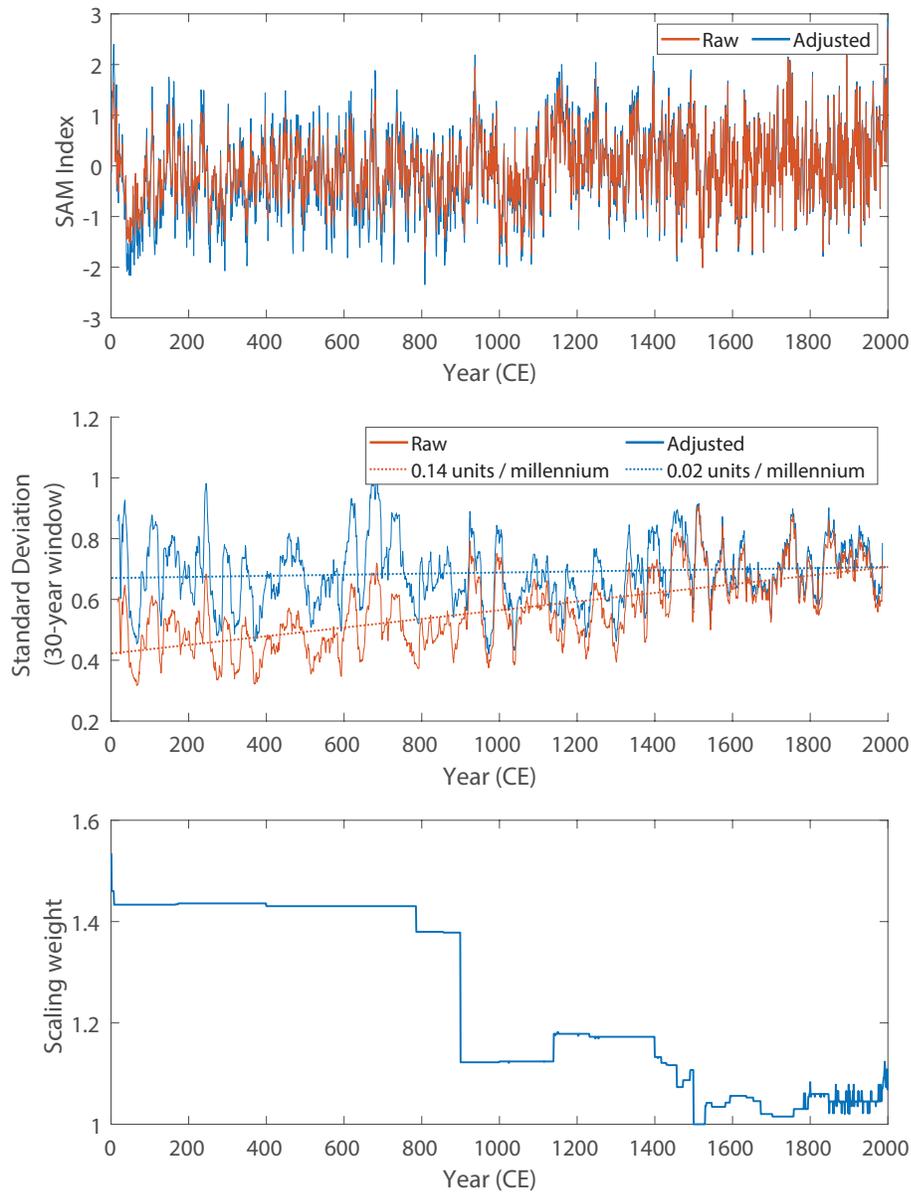


Figure B.9: Variance correction scheme. Top: Raw assimilation (red) and the variance-adjusted final reconstruction (blue). The raw assimilation displays less variance than the final reconstruction, particularly during the early part of the record. Middle: Solid lines show the moving 30-year standard deviations of the raw assimilation (red) and variance-adjusted reconstruction (blue). Dotted lines indicate the linear trends in the moving standard deviation time series over the full period. Bottom: The multiplicative scaling weights used to produce the final, variance-adjusted reconstruction. Weights are applied multiplicatively to the raw assimilation in each time step.

	Climate Model	Acronym	Output Resolution (Latitude x Longitude)	Experiments	Number of Years
High-Resolution	CCSM4	CCSM4	0.94° x 1.25°	past1000, historical	1156
	CESM1.1-CAM5	CESM	1.89° x 2.50°	LME full-forcing 2	1156
	MPI-ESM-P	MPI	1.86° x 1.88 °	past1000, historical	1156
	MRI-CGCM3	MRI	1.12° x 1.13°	past1000, historical	1156
	BCC-CSM1-1	BCC	2.79° x 2.81°	past1000	1151
	CSIRO-Mk3L-1-2	CSIRO	3.18° x 5.63°	past1000, historical	1150
	FGOALS-g1	FGOALS	4.62° x 5.00°	past1000	1000
	HadCM3	HadCM3	2.50° x 3.75°	past1000, historical	1147
	IPSL-CM5A-LR	IPSL	1.89° x 3.75°	past1000, historical	1156
	MIROC-ESM	MIROC	2.79° x 2.81°	past1000, historical	1156

Table B.2: Climate models tested for use in assimilation priors.

Model	SAM Latitudes	Correlation
BCC	43°S, 66°S	-0.90
CCSM4	47°S, 69°S	-0.79
CESM	46°S, 69°S	-0.89
CSIRO	43°S, 65°S	-0.90
HadCM3	45°S, 68°S	-0.84
FGOALS	31°S, 59°S	-0.84
IPSL	39°S, 63°S	-0.94
MIROC	40°S, 63°S	-0.90
MPI	42°S, 66°S	-0.83
MRI	44°S, 69°S	-0.90
Gong et al. (1999)	40°S, 65°S	

Table B.3: Latitudes with the most strongly anticorrelated zonal-mean SLP anomalies in tested climate models. Correlation coefficients are calculated for DJF seasonal means over all available years in the interval 850-2005 CE.

Proxy Type	PAGES2k ID	Site Name	Latitude, Longitude	Years (CE)	Season
Coral	Ocn 114	Houtman Abrolhos	28.47°S, 113.77°E	1795 - 1994	July - June
	Ocn 153	Houtman Abrolhos	28.46°S, 113.77°E	1798 - 2000	January - December
	Ocn 154	Houtman Abrolhos	28.46°S, 113.75°E	1848 - 2000	January - December
	Ocn 155	Houtman Abrolhos	28.46°S, 113.75°E	1848 - 2000	January - December
	Ocn 158	Houtman Abrolhos Islands	28.47°S, 113.77°E	1900 - 2000	January - December
Glacier Ice	Ant 001	Talos Dome	72.80°S, 159.06°E	1232 - 1995	January - December
	Ant 002	DSS	66.77°S, 112.81°E	173 - 1995	January - December
	Ant 003	Plateau Remote	84.00°S, 43.00°E	2 - 1986	January - December
	Ant 004	Coastal DML	70.86°S, 11.54°E	1533 - 1994	January - December
	Ant 005	Site DML05	75.00°S, 0.01°W	166 - 1996	January - December
	Ant 006	WDC05A	79.46°S, 112.09°W	786 - 2000	January - December
	Ant 007	WDC06A	79.46°S, 112.09°W	1 - 2000	July - June
	Ant 008	US-ITASE-2000-1	79.38°S, 111.24°W	1673 - 2000	July - June
	Ant 010	James Ross Island	64.20°S, 57.68°W	1 - 2000	January - December
	Ant 011	Siple Station	75.92°S, 84.25°W	1417 - 1983	January - December
	Ant 012	Berkner Island (South)	79.57°S, 45.72°W	1000 - 1992	January - December
	Ant 017	Ferrigno	74.57°S, 86.90°W	1703 - 2000	January - December
	Ant 019	MES	77.52°S, 167.68°E	1473 - 2000	January - December

*Continued on next page*

*Continued from previous page*

Proxy Type	PAGES2k ID	Site Name	Latitude, Longitude	Years (CE)	Season
Glacier Ice	Ant 020	Site DML07	75.58°S, 3.43°W	1000 - 1994	January - December
	Ant 021	Site DML17	75.17°S, 6.50°E	1000 - 1997	January - December
	Ant 024	US-ITASE-2002-4	86.50°S, 107.99°W	1594 - 2000	July - June
	Ant 025	VLG	77.33°S, 162.53°E	1140 - 2000	January - December
	Ant 026	Vostok	78.28°S, 104.80°E	1654 - 2000	January - December
	Ant 028	WDC06A	79.46°S, 112.09°W	1 - 2000	July - June
Borehole	Ant 027	WAIS-Divide	79.46°S, 112.12°W	8 - 2000	January - December
Lake Sediment	SAm 003	Laguna Aculeo	33.85°S, 70.92°W	856 - 1997	December - February
	SAm 030	Laguna Chepical	32.27°S, 70.50°W	1 - 2000	November - February
	SAm 031	Laguna Escondida	45.52°S, 71.82°W	400 - 2000	January - December
Tree	Aus 001	Mt. Read	41.83°S, 145.53°E	1 - 2000	November - April
	Aus 002	Oroko	43.23°S, 170.28°E	900 - 1999	September - April
	Aus 004	CTP East Tasmania	41.31°S, 147.75°E	1430 - 1994	September - November
	Aus 005	Pink Pine NZ	43.00°S, 171.00°E	1457 - 1999	September - April
	Aus 007	Buckleys Chance Tasmania	42.27°S, 145.87°E	1463 - 1991	October - April
	Aus 009	CTP West Tasmania	41.67°S, 145.65°E	1547 - 1998	June - August

*Continued on next page*

*Continued from previous page*

Proxy Type	PAGES2k ID	Site Name	Latitude, Longitude	Years (CE)	Season
Tree	Aus 030	Stewart Island	47.00°S, 167.80°E	1758 - 1993	September - April
	Aus 031	Takapari Cedar	40.07°S, 175.98°E	1530 - 1992	September - April
	SAm 006	Central Andes composite 11	40.10°S, 72.05°W	1492 - 1995	May - April
	SAm 024	Central Andes composite 6	38.50°S, 71.50°W	1435 - 2000	May - April
	SAm 025	Central Andes composite 9	39.33°S, 71.25°W	1636 - 2000	May - April
	SAm 029	Central Andes composite 15	41.17°S, 71.92°W	1582 - 1991	May - April

Table B.4: PAGES2k Sites used in the final reconstruction. The included sites are south of 25°S, and have annual or higher resolution.

## APPENDIX C

## DASH: A MATLAB toolbox for paleoclimate data assimilation

King, J., Tierney, J., Osman, M., and Anchukaitis, K. DASH: A MATLAB toolbox for paleoclimate data assimilation. *prepared for submission to Geoscientific Model Development*.

## **DASH: A MATLAB toolbox for paleoclimate data assimilation**

Jonathan King<sup>1,2</sup>, Jessica Tierney<sup>1</sup>, Matt Osman<sup>1</sup>, Kevin Anchukaitis<sup>1,2,3</sup>

<sup>1</sup> Department of Geosciences, University of Arizona, Tucson, Arizona, USA

<sup>2</sup> Laboratory of Tree-Ring Research, University of Arizona, Tucson, Arizona, USA

<sup>3</sup> School of Geography, Development, and Environment, University of Arizona, Tucson, Arizona, USA

### **C.1 Abstract**

Paleoclimate data assimilation (DA) is a novel tool for reconstructing past climates that directly integrates proxy records with climate model output. Despite the potential for DA to expand the scope of quantitative paleoclimatology, these methods remain difficult to implement in practice due to the multi-faceted requirements and data handling needed for DA reconstructions, the diversity of DA methods, and the need for computationally efficient algorithms. Here, we present DASH, a MATLAB toolbox designed to facilitate paleoclimate DA analyses. DASH provides command line and scripting tools that implement the common tasks in DA workflows. The toolbox is highly modular and is not built around any specific analysis, and thus DASH supports paleoclimate DA for a wide variety of time periods, spatial regions, proxy networks, and algorithms. DASH includes tools for integrating and preparing data stored in disparate formats, building state vector ensembles, running proxy (system) forward models, and implementing various DA algorithms. The toolbox also provides optimized algorithms for implementing ensemble Kalman filters, particle filters, and optimal sensor analyses with variable and modular parameters. This paper reviews the key components of the DASH toolbox and presents examples illustrating DASH's use for palaeoclimate DA applications.

## C.2 Introduction

Past climates provide insight into the drivers, variability, and evolution of the Earth's climate system, and are invaluable for providing insight on the consequences of current and future anthropogenic climate change (Alley, 2003; Hargreaves et al., 2007; Rice et al., 2009; Schmidt, 2010; Snyder, 2010; Bell et al., 2013; Ault et al., 2014; Coats et al., 2020; Tierney et al., 2020a). Paleoclimate studies can help constrain important climate system properties including Equilibrium Climate Sensitivity (Hegerl et al., 2006; Rohling et al., 2012; Hansen et al., 2013; Kutzbach et al., 2013; Rohling et al., 2018; Sherwood et al., 2020; Tierney et al., 2020b; Zhu et al., 2020b), quantify internal and forced variability across a range of timescales and climate system metrics (Cane et al., 2006; Cook et al., 2011; Goosse et al., 2012a; Ault et al., 2013; Fernández-Donado et al., 2013; Neukom et al., 2019; Fang et al., 2021), and can serve as analogues for future warm climate states projected to occur due to anthropogenic warming (Overpeck et al., 2006; Burke et al., 2018; Tierney et al., 2020a). Reconstructions of past climates also provide out-of-sample targets used to assess the skill of climate models, which in turn helps constrain future projections and enables superior climate change adaptation strategies (Crowley, 1991; Hargreaves and Annan, 2009; Schmidt et al., 2013; Zhu et al., 2021a,b; Gulev et al., 2021).

Beyond the limited period of instrumental climate observations, researchers have primarily relied on two methods for studying past climates: proxy reconstructions and climate model hindcasts. In a proxy reconstruction, paleoclimatologists use climate proxy records, such as tree rings, ice cores, speleothems, corals, and lake and marine sediments, to make statistical estimates of past climates. These reconstructions rely on a combination of empirical and process-based understanding to link proxy records to features and characteristics of the Earth's climate system. By contrast, climate model hindcasts leverage general circulation models to simulate past climate states using estimates of past boundary conditions, such as the Earth's orbital parameters, atmospheric greenhouse gas concentrations, volcanic eruptions, continental configurations, and land cover. Both methods for studying past climate have strengths and weaknesses. A major advantage of using

proxy data to reconstruct past climates is that they produce estimates of temperature, precipitation, or other climate variables that are consistent with the actual trajectory of the Earth's climate system. These reconstructions can also provide independent validation of climate model performance. However, many factors can hinder the inference of past climates from proxy data. These include the sparse distribution of proxy records through space and time, time-uncertainty due to limits on the precision of geochronology, and the presence of non-climatic noise within proxy records. Furthermore, the physical processes that archive climate signals in proxy records can be complex and are often not completely understood, which complicates the extraction of climate signals from proxy data using linear, univariate, and empirical statistical approaches. Proxy records are sensitive to the local climates in which they form, but many reconstructions target large-scale climate features or ocean-atmosphere modes not directly sensed by the available proxy data. Some reconstructions derive relationships between proxy records and target variables using calibrations with the instrumental era; however, modern climate is not in equilibrium and continues to respond to increasing anthropogenic climate forcings. Therefore, modern teleconnections and climate-system spatial covariances may differ from long-term and unforced patterns. Finally, many proxy reconstruction methods assume that teleconnections between local- and large-scale climate variables are stationary over reconstruction periods, an assumption that may not hold in reality.

By contrast with proxy reconstructions, climate model hindcasts simulate data for target climate variables at all spatial points and time steps within the model domain. Furthermore, these simulated climate variables evolve according to fundamental physical governing equations and parameterizations, rather than the statistical associations and assumptions typically used for proxy reconstructions. Consequently, paleoclimate simulations can provide insight into the physical mechanisms behind reconstructed climate phenomena. However, no model fully captures the real Earth system, and so all paleoclimate hindcasts necessarily contain errors in their representation of past climates. Additionally, many model variables are dominated by internal variability, sensitivity to initial conditions, and/or chaotic behavior over a range of time periods (Deser et al., 2012). Thus, no individual simulation captures the true or specific trajectory of the Earth's past

climate; instead, each simulation represents a single possible trajectory in a distribution of physically-plausible past climate states (e.g. Kay et al., 2015). Finally, climate models require external validation to evaluate their fidelity and accuracy in reproducing past climate states.

Recently, data assimilation (DA) methods have emerged as a novel approach to the problems and challenges of paleoclimate reconstruction (e.g. Bhend et al., 2012; Goosse et al., 2012b; Mairesse et al., 2013; Hakim et al., 2016; Steiger et al., 2017, 2018; Tardif et al., 2019; Tierney et al., 2020b; King et al., 2021; Osman et al., 2021; Zhu et al., 2022; King et al., 2022). Unlike the two independent approaches described above, DA methods integrate proxy data directly with climate model output and thereby leverage the strengths of both information sources. By leveraging climate model simulations, DA methods provide are able to provide global reconstructions (e.g. Evans et al., 2001), and the relationships between all simulated variables are linked through the physically based governing equations of the model. Simultaneously, DA reconstructions are constrained by proxy records and thus reflect the true trajectory of the Earth's past climate. DA methods use forward models to describe how climate signals are sensed by and recorded in proxy archives, and thus can incorporate proxy system physical processes that are multivariate or nonlinear. Furthermore, the use of proxy forward models allows DA methods to relax calibration requirements when attempting to reconstruct large-scale climate modes or fields, such that proxy data can be calibrated to local climate variables rather than directly to large-scale teleconnections. DA methods also relax assumptions of teleconnection stationarity, as the effects of changing climate boundary conditions can be reflected in the evolution of climate model output and its covariance.

Despite the potential strengths of paleoclimate DA, these reconstructions are difficult to implement in practice. DA analyses must perform a number of tasks, including integrating the output from climate model simulations, proxy records, and possibly instrumental data, all of which may use different data formats, units, and metadata. The number of potential reconstruction targets and proxy variables is immense, and the choice of parameters will affect the implementation of any particular DA reconstruction (compare Tardif et al., 2019; Tierney et al., 2020b; King et al., 2021; Osman et al., 2021; King

et al., 2022). Consequently, it can be difficult to adapt codes implementing an existing reconstruction to alternative applications. Paleoclimate DA also encompasses a diverse array of algorithms and algorithm variants (compare Goosse et al., 2006; Dubinkina and Goosse, 2013; Steiger et al., 2014; Matsikaris et al., 2015; Comboul et al., 2015; Dee et al., 2016; Acevedo et al., 2017; Liu et al., 2017; Perkins and Hakim, 2017; Franke et al., 2020), further increasing the complexity of implementing DA codes. Finally, DA methods are often computationally intensive and require both optimized algorithms and efficient use of computer memory, and these considerations can dissuade potential users lacking experience with or access to high-performance computing.

Although DA software does exist, thus far these packages are not suitable for generalized paleoclimate applications with a diverse range of time scales, climate model requirements, and proxy data types. Packages designed to implement general DA methods typically lack support for fundamental components of paleoclimate DA, such as the use of proxy forward models. By contrast, DA packages designed for paleoclimate applications, such as the LMR (Hakim et al., 2016; Tardif et al., 2019) or PHYDA (Steiger et al., 2018), have been built to implement specific analyses, proxy data, and climate model inputs. Adapting these products for generalized paleoclimate applications requires modifying the source code, which may be difficult and time-intensive and thus presents a barrier to their use.

A second difficulty for paleoclimatologists seeking to implement DA is that the methods are comparatively complex relative to existing reconstruction methods. Because of their multi-faceted natures, describing experimental DA setups in sufficient detail to allow reproducibility requires considerable length, and published methods may focus on the broad scope of the mathematics while neglecting key implementation steps in favor of brevity. Additionally, there are still relatively few paleoclimate applications in the mathematical DA literature, so DA descriptions may use a variety of inconsistent mathematical notations. Finally, the diversity of algorithm variants further hinders transparency and accessibility, as studies using similarly named algorithms may implement different methods in practice (compare Tardif et al., 2019; Franke et al., 2020; Tierney et al., 2020b; King et al., 2022). Ultimately, there are limited frameworks for discussing DA within

the paleoclimate literature and the field as a whole would benefit from a more transparent implementations that do not require additional specialized training.

In this paper, we present DASH, a MATLAB toolbox supporting paleoclimate data assimilation. The toolbox is designed for general paleoclimate DA and is not built around or for any particular analysis, time period, proxy type, or climate model. Consequently, the toolbox is highly modular and allows flexible implementation of diverse DA analyses. DASH provides command line / scripting utilities designed to implement common tasks for paleoclimate DA workflows, with a goal of improving access to DA methods for users with diverse scientific backgrounds. DASH includes support for organizing climate data, building state vector ensembles, running common proxy forward models, and implementing common DA algorithms. All algorithms are optimized for both speed and efficient-memory use. An additional goal of DASH is to improve clarity in DA analyses and provide a framework for paleoclimate DA discussions. Consequently, DASH commands are written in a style supporting human-readable analytical codes.

The remainder of this paper is organized as follows. In Section 2, we present a brief overview of paleoclimate DA, with the aim of introducing common tasks, data, and algorithms for paleoclimate DA workflows. In Section 3, we describe the DASH toolbox specifically. We detail its general characteristics and layout and then describe its major components. In Section 4, we provide two examples that use DASH to implement paleoclimate data assimilation. These examples use diverse temporal periods, spatial regions, proxy networks, and algorithms in order to demonstrate the flexibility of the DASH toolbox. Finally, we provide a set of best practices and caveats, discuss the DASH toolbox in the broader context of paleoclimate DA, and describe potential and anticipated future developments to the code.

### **C.3 Overview of Paleoclimate DA**

In this section, we provide a brief overview of paleoclimate data assimilation. The goal of this section is to introduce DA for paleoclimate researchers who may not be familiar with the broader mathematical DA literature. In particular, our aims are to (1) Provide

accessible insight into the DA “black-box”, (2) improve the transparency of common DA algorithms, (3) establish a vocabulary for DA workflows, and (4) provide context for the DASH software package. We focus on illustrating tasks and quantities common in paleoclimate DA workflows, rather than providing complete mathematical descriptions (which can be found elsewhere, e.g. Evensen (1994); Van Leeuwen (2009)). Here, we focus specifically on the ensemble Kalman filter and ensemble particle filter methods. We also describe an optimal sensor algorithm based on an ensemble Kalman Filter framework. Additional and more complete descriptions of DA algorithms are available in Evensen (1994); Anderson and Anderson (1999); Whitaker and Hamill (2002); Goosse et al. (2006, 2012b); Dubinkina and Goosse (2013); Steiger et al. (2014); Comboul et al. (2015); Hakim et al. (2016); Tardif et al. (2019); Franke et al. (2020); Tierney et al. (2020b); King et al. (2021); Osman et al. (2021)

Conceptually and in the broadest terms, DA methods combine output from climate model simulations ( $X_p$ ) with proxy records ( $Y$ ) to reconstruct a set of target climate variables ( $X_a$ ).

$$X_a = f(X_p, Y) \quad (\text{C.1})$$

The reconstruction proceeds by updating climate variables from the climate models  $X_p$  to more closely match the proxy records  $Y$ . The updated climate variables  $X_a$ , also known as the analysis, thus form the reconstruction. The Kalman filter and particle filter methods discussed in this paper can also be formulated as Bayesian filters (Chen et al., 2003; Wikle and Berliner, 2007) wherein new information ( $Y$ ) is used to update estimates of state parameters ( $X$ ). Hence, we will often refer to  $X_p$  and  $X_a$  as the *prior* and *posterior*, respectively.

In general, climate model output is organized into *state vectors*, which consist of multi-dimensional spatiotemporal climate model output reshaped into a vector of data values (Figure C.1, upper left). There is no strict definition for the contents of a state vector, but they typically include data for one or more climate variables at a set of spatial points. A state vector might also contain a trajectory of successive points in time; for example, individual months of the year or a number of successive years following an event of interest. Essentially, a state vector serves as a possible description of the climate

system for some period of time. In this paper, we focus on *ensemble* DA methods, which rely on state vector ensembles. A *state vector ensemble* is a collection of multiple state vectors organized in a matrix (Figure C.1, upper right), and a given ensemble provides an empirical distribution of possible climate states. For paleoclimate applications, ensemble members are often selected from different points in time, different members of a model ensemble, or both. In a typical DA algorithm, the state vectors in an ensemble are compared to a set of proxy record values in a given time step. In other words, the method compares potential descriptions of the climate system taken from the climate model to the proxy values from the real past climate record. The similarity of each state vector to the set of proxy records is then used to inform the final reconstruction.

In order to compare state vectors with a set of proxy record values, DA methods must transfer state vectors and proxy records into a common unit space. This is accomplished by applying proxy forward models (Evans et al., 2013) to relevant climate variables stored in each state vector (Figure C.1, bottom left). Applying a forward model to a state vector produces a value in the same units as the corresponding proxy record and therefore allows direct comparison of the state vector and observed proxy value. In general, DA methods will run a forward model to estimate each proxy record for each state vector in an ensemble; the collective outputs are referred to as *proxy estimates* ( $\hat{Y}$ ) and allow comparison of the states in the ensemble with a set of proxy records. The difference between the proxy observations and proxy estimates is known as the *innovation* (Figure C.1, bottom right):

$$\text{innovation} = Y - \hat{Y} \quad (\text{C.2})$$

and describes the discrepancies between the actual proxy records and the climate states in the ensemble. The innovation is then used to constrain or update the prior ensemble ( $X_p$ ) to more closely resemble the observed proxy records.

In addition to proxy innovations, the DA methods detailed here also consider proxy uncertainties ( $R$ ) when comparing state vectors to the proxy records, such that:

$$X_a = f(X_p, Y, R) \quad (\text{C.3})$$

In this way, proxy records with high uncertainties are given less weight in the reconstruction. In classical assimilation frameworks, proxy (observation) uncertainties ( $R$ ) are often

derived from the uncertainty inherent in measuring an observed quantity. For example, the uncertainty of width measurements in a tree-ring chronology. However, in nearly all paleoclimate applications, measurement uncertainties are small compared to (1) the uncertainties inherent in proxy forward models, and (2) uncertainties resulting from non-climatic noise in the proxy records. Thus, in paleoclimate DA the proxy uncertainties  $R$  must account for proxy noise and forward model errors, as well as the covariance between different proxy uncertainties. Most generally,  $R$  is the proxy error-covariance matrix. This matrix is diagonal when proxy errors are assumed uncorrelated; otherwise,  $R$  is a full covariance matrix.

When using a Kalman Filter (Kalman, 1960; Andrews, 1968; Evensen, 1994), the update equation is given by:

$$X_a = X_p + K(Y - \hat{Y}) \quad (\text{C.4})$$

Equation C.4 shows that the innovation is weighted by the Kalman Gain matrix ( $K$ ) in order to compute an update for each state vector in the prior ensemble ( $X_p$ ). The Kalman Gain weighting considers multiple factors, including (1) the covariance of the proxy estimates ( $\hat{Y}$ ) with target climate variables ( $X_p$ ), (2) the covariance between the proxy estimates ( $\hat{Y}$ ), and (3) the uncertainties in the proxies ( $R$ ), such that:

$$K = \text{cov}(X, \hat{Y})[\text{cov}(\hat{Y}) + R]^{-1} \quad (\text{C.5})$$

Applying the updates produces an updated (posterior) ensemble ( $X_a$ ), such that the climate states in  $X_a$  will more closely resemble those recorded by the real proxy records ( $Y$ ). The ensemble nature of  $X_a$  is also advantageous because the distribution of climate variables across  $X_a$  can help quantify the uncertainty in the reconstruction.

By contrast with Kalman filters, particle filters (Van Leeuwen, 2009) combine the innovation with proxy record uncertainties ( $R$ ) to compute a weight for each state vector in the ensemble. The reconstruction is then calculated as a weighted mean of the state vectors in the ensemble. Classical particle filters compute these weights using a Bayesian scheme, such that each state vector  $i$  is first assigned an importance weight:

$$s_i = \exp\left[-\frac{1}{2} (Y - \hat{Y}_i)^T R (Y - \hat{Y}_i)\right] \quad (\text{C.6})$$

and then importance weights are normalized to give the final state vector weights:

$$w_i = \frac{s_i}{\sum_{j=1}^N s_j} \quad (\text{C.7})$$

However, classical particle filters can suffer from degeneracy in the high-dimensional systems common to paleoclimate DA. Essentially, a single ensemble member receives a weight of 1, whereas all other ensemble members receive near-zero weights. When this occurs, reconstructed values ( $X_a$ ) resemble the single state vector most similar to the proxy records, rather than values across an ensemble. A common correction for degeneracy involves using the mean of the  $N$  state vectors with the highest Bayesian weights. By contrast, the “degenerate particle filter” refers to the case when the single best state vector is used as the reconstruction (e.g. Goosse et al., 2006, 2010). The “analogue method” may also refer to a degenerate particle filter (e.g. Goosse et al., 2006), although the meaning of this term varies throughout the paleoclimate literature.

The optimal sensor algorithm described in this paper follows the method presented by Comboul et al. (2015). This method is derived from an ensemble Kalman filter and complements the reconstruction framework by providing additional information about the contribution of proxy data sites to the reconstruction. In paleoclimate, optimal sensor analyses have traditionally been used to assess the proxy network necessary to skillfully reconstruct a climate field, to evaluate the potential of new proxy sites, and to prioritize future proxy development (e.g. Bradley, 1996; Evans et al., 1998; Comboul et al., 2015). Here, we expand the method to assess the relative influence of individual proxy records on a reconstructed index. Rather than reconstructing climate variables over time, the algorithm instead tests the ability of a proxy record to reduce the variance of a climate metric  $J$  across an ensemble. A proxy record’s ability to reduce variance is determined using the covariance of its estimates ( $\hat{y}$ ) with the climate metric ( $J$ ) combined with the uncertainty of the proxy record ( $R$ ). For a given proxy record ( $y$ ), this equation is given by:

$$\Delta\sigma_k = \text{cov}(\hat{y}, J)^2 [\text{var}(\hat{y}) + R]^{-1} \quad (\text{C.8})$$

and the proxy that most strongly reduces variance is selected as the optimal sensor:

$$s_{\text{optimal}} = \text{argmax } \Delta\sigma \quad (\text{C.9})$$

This proxy is used to update the climate metric using an ensemble Kalman Filter (Equations C.4, C.5) and then removed from the network. This analysis then iterates using the remaining sensors until the desired number of sensors are selected. Ultimately, the method both ranks the proxies in a network and also assesses the total variance reduced by a particular proxy network. This method requires proxy estimates ( $\hat{Y}$ ) to calculate climate metric covariance but does not use proxy record values themselves ( $Y$ ), as the potential to reduce ensemble variance is independent of actual proxy values.

## C.4 Description of DASH

### C.4.1 General Characteristics

DASH is a MATLAB toolbox designed to help implement paleoclimate data assimilation. The code is designed for use from the command line as well as within scripts and functions. DASH is written in an object-oriented style, which supports the modularity of the code; the toolbox consists of several classes and packages, each implementing a common task for paleoclimate DA. The code is intended for users with basic previous experience with MATLAB; in particular, users will benefit from knowing how to write a basic `for` loop, and how to index into arrays.

A stated goal of the DASH toolbox is to support the transparency of paleoclimate data assimilation analyses, and the object-oriented design supports this aim. DASH methods are accessed via dot-indexing, which improves clarity by placing sub-tasks within the context of a larger piece of the data assimilation process. Additionally, tasks with many parameters or options are organized into objects, which can store settings between commands. Consequently, the parameters used to implement a task are split across several commands, improving both the clarity and modularity of codes utilizing DASH.

To support command-line workflows, DASH is designed for console display and does not rely on a graphical user interface (GUI). Users can inspect the state of class objects, assimilation analyses, and other DASH components by displaying them in the console. Users can also examine reference guides for DASH components using the `help` command; however, we recommend that users instead use the HTML documentation set,

which is detailed below. DASH anticipates that users may not be familiar with all aspects of paleoclimate data assimilation, or with all components of the toolbox. DASH therefore implements robust input checking and error handling for all user-facing methods. Error messages are designed to clearly communicate input failures and suggest possible solutions without requiring users to know the inner workings of the DASH codebase.

To install DASH, users should first download a stable release of the toolbox, which can be found at the project's Github repository (<https://github.com/JonKing93/DASH/releases>), MATLAB FileExchange ([reserve-url-pending-review](#)), or in the MATLAB Add-On Explorer. Then, open the downloaded DASH.mltbx file to complete the installation. We encourage users to download one of the project's stable releases, as the source code on the Github repository's main branch may be in active development and is not configured for quick installation.

The DASH toolbox is accompanied by comprehensive documentation written in HTML. This documentation includes (1) a reference guide for every class, package, method, and function, (2) example use cases, and (3) How-Tos and FAQs for common tasks and troubleshooting. The entire documentation can be accessed by entering the `dash.doc` command from the MATLAB command line. Alternatively, users can open the reference manual for a particular component by providing the component name as input: `>> dash.doc("component name")`. The documentation is also available on the project's website (<https://jonking93.github.io/DASH>).

#### **C.4.2 DASH Components**

DASH consists of several classes and packages, each implementing a particular task commonly required for paleoclimate data assimilation (Figure C.2). In brief, the toolbox contains components to (1) organize and catalogue input data, (2) design and build state vector ensembles, (3) estimate proxy records via proxy forward models, and (4) implement common data assimilation algorithms. In the remainder of this section, we examine the characteristics and features of each of these modules.

**Organize Climate Data: `gridfile`, `gridMetadata`**

We begin our overview with the `gridfile` class. This module is intended to facilitate combining datasets stored in different formats and with disparate metadata. `gridfile` associates the data values with user-specified metadata, which allows users to manipulate large datasets using their preferred and human-readable metadata. The primary purposes of the class include consolidating datasets split across multiple files, promoting human-readable data manipulation, and uniting disparate data formats within an intuitive framework. The class implements `gridfile` objects, which act as a catalogue for the data stored in various source files. Each `gridfile` catalogue manages an abstract  $N$ -dimensional grid, whose scope is defined by user-provided dimensional metadata. This allows users to catalogue datasets of varying dimensionality, while simultaneously tagging data elements with unique and user-preferred metadata values. We note that the grid abstraction does not imply that `gridfile` datasets must use a Cartesian spatial grid. Rather, the class supports a wide-variety of spatial layouts, including rectilinear systems, tripolar grids, randomly distributed spatial sites, and datasets without any spatial component at all.

After first defining the scope of a `gridfile`, users can add the data source files to the catalogue by associating the data in each file with a portion of the  $N$ -dimensional grid. In this way, the data in each source file is placed within the context of the overall dataset. The `gridfile` package supports data source file formats common in paleoclimate DA – including NetCDF, OPeNDAP, MATLAB’s binary MAT files, and delimited-text files – and individual catalogues may contain any mixture of file formats. The contents of each catalogue are saved in a `.grid` file, so data catalogues can persist across multiple coding sessions. We emphasize that these `.grid` files save only a *catalogue* of a dataset, and not the dataset itself. Thus, `.grid` files do not duplicate data, and individual `.grid` files remain small (typically a few kilobytes) even when they refer to datasets spanning many gigabytes of memory. Once a catalogue is complete, users can return data using the `load` command, which provides a common interface for accessing data in the catalogue. Users can also return a subset of the catalogued data by querying the associated metadata.

The `gridfile` class also allows users to apply data transformations, such as log transforms or fill values, to a catalogue. Such transformations are only applied to loaded data, which improves computational efficiency and maintains the data sources as read-only files. Finally, the class allows users to perform arithmetic operations across multiple gridfile datasets; these operations are analogous to several commonly used NetCDF operators but are not limited to NetCDF files.

The `gridMetadata` class implements the object used to define the metadata for a dataset. This class plays a relatively minor role within the DASH toolbox and it is mainly used to define the scope of `gridfile` datasets and to locate data subsets within a `gridfile` catalogue. We contrast `gridMetadata` with `ensembleMetadata`, a second metadata class implemented by DASH. Whereas `gridMetadata` characterizes values in an  $N$ -dimensional dataset, `ensembleMetadata` instead characterizes  $N$ -dimensional datasets after they are reshaped into state vector ensembles. Further details for the `ensembleMetadata` class are given in Section C.4.2.

### **Build state vector ensembles: `stateVector`, `ensemble`**

The next key component of DASH is the `stateVector` class. This component is designed to facilitate flexible design of state vector ensembles while minimizing the amount of data manipulation done by the user. The class implements objects that hold design parameters required to build a state vector ensemble from `gridfile` catalogues. To design a state vector, users first initialize a `stateVector` object and then initialize variables in the design. Each variable is associated with a `gridfile` dataset, and multiple variables in the state vector may be derived from the same dataset. We note that when a user adds a variable to a `stateVector` object, no data is loaded into memory at that time. Instead, the object initializes a set of design parameters that can later be used to extract data for the variable from its `gridfile`. To design the state vector, users next specify options for the dimensions of the variables. As a first step, users should indicate which dataset dimensions are used to select ensemble members. In most paleoclimate DA applications, ensemble members are selected from different time steps and/or different climate model simulations. However, `stateVector` is highly flexible and also allows ensembles built

along other dimensions; for example, ensembles built from different height levels or from different spatial locations and sites. Users can also specify a subset of elements along an ensemble dimension to use for building ensemble members. For example, in a dataset with monthly resolution, a user could specify to only select ensemble members from January time steps. The class provides many additional methods for designing state vector variables: users can specify that a variable should use a subset of a `gridfile` dataset, compute means, weighted means, or principal components over data dimensions, and select options for processing variables with different metadata formats. Users can also specify that individual ensemble members should contain temporal sequences. For example, a variable could include data from individual months of the year, useful for seasonal analyses, or from successive years, which supports superposed-epoch analyses for climate conditions following discrete events of interest.

Once a design is complete, users call the `build` command, which loads necessary data from the `gridfile` catalogues and builds a state vector ensemble according to the specified design parameters. When building a state vector ensemble, `stateVector` will ensure that all variables within a given ensemble member align to the same metadata values. For example, in an ensemble selected from different time steps, the data for the variables in each ensemble member will all correspond to the same time step. Similarly, in an ensemble selected from different model simulations, the variables in each ensemble member will all be drawn from the same simulation. The class also ensures that ensemble members are constructed from complete data. For example, if a state vector variable includes a temporal mean or sequence, then the `build` method will never select an ensemble member for which the mean or sequence would extend outside of the dimensions of the dataset.

When building an ensemble, users have the option to return the ensemble directly as an array, or to save the ensemble to a file. This later option is useful, as state vector ensembles may exceed the size of active memory, particularly when state vectors include multiple spatial fields from high-resolution climate models. In the DASH framework, these files are saved with a `.ens` extension, and the toolbox provides the `ensemble` class to facilitate memory-efficient interactions with saved state vector ensembles. We

highlight the ability of the `ensemble` class to selectively load requested state vector rows, variables, and ensemble members into memory. These features have particular utility when running (1) proxy forward models, which typically only require a small subset of ensemble data, and (2) data assimilation algorithms, as many reconstructions only target a subset of variables in an ensemble. Users can also call the `evolving` command to implement evolving offline priors (e.g. Osman et al., 2021) without loading data values to memory.

### **Proxy Forward Models: `PSM`, `ensembleMetadata`**

After building a state vector ensemble, a common next task in paleoclimate DA is to design a forward model for each proxy record. These forward models are either used to generate proxy estimates (for offline assimilations) or provided directly as input to data assimilation algorithms (for online regimes). The `PSM` package facilitates all these tasks by providing users modular access to commonly used proxy system forward models. The actual implementation of proxy system models is beyond the scope of DASH; instead, the `PSM` package unites required interactions with available proxy model codes under a common interface. DASH currently supports multivariate linear models (see Hakim et al., 2016; Tardif et al., 2019; Zhu et al., 2020a), the Vaganov-Shashkin ‘Lite’ (VSL) tree ring model (Tolwinski-Ward et al., 2011), the `BAY*` suite of Bayesian foraminiferal and membrane-lipid models Tierney and Tingley (2014); Malevich et al. (2019); Tierney et al. (2019); Tierney and Tingley (2018), a Palmer Drought Severity Index (PDSI) estimator Guttman (1991); Van der Schrier et al. (2011), and the models within the `PRYSM` Python package (Dee et al., 2015) (Table 1). We anticipate that this list will grow with future advances in proxy system modeling.

Users can call the `download` method to automatically download selected models from their respective Github repositories and add them to the MATLAB active path. The class then allows users to design `PSM` objects, which implement a forward model for a particular proxy record with modular model parameters. Users then indicate which state vector rows hold the data needed to run each forward model; this search is facilitated by the `ensembleMetadata` class detailed in the next paragraph. Users can then either

use the `estimate` command to run the forward models over the state vector ensemble and generate proxy estimates, or provide the completed forward models directly to an online DA algorithm. Users can also run the forward models over assimilated state vector ensembles, in order to validate proxy records against assimilation results (e.g. Tardif et al., 2019; Tierney et al., 2020b; King et al., 2021; Osman et al., 2021).

The process of running forward models on a state vector ensemble is facilitated by the `ensembleMetadata` class. This class implements objects that organize metadata along the rows and columns of a state vector ensemble. An `ensembleMetadata` object is created whenever a user builds a state vector ensemble and can also be returned for `.ens` files and `stateVector` objects. The class can be used to locate state vector rows corresponding to particular variables, spatial locations, or time sequences, and can also be used to locate specific ensemble members. A major task of `ensembleMetadata` is to locate state vector rows that correspond to proxy forward model inputs. In addition to locating specific variables, the class can determine which data elements are closest to the location of a proxy site, which is frequently useful when extracting forward model inputs from large climate model fields. Each `ensembleMetadata` object also holds the metadata necessary to reshape state vectors back into gridded datasets. Consequently, the class is also used to reshape DA outputs back into spatial grids for post-processing and visualization.

### **Data Assimilation Algorithms: `kalmanFilter`, `particleFilter`, `optimalSensor`**

This section describes the classes used to implement data assimilation algorithms. Each class implements objects that hold parameters for a particular type of analysis. The object-oriented layout allows users to specify diverse algorithm parameters, while promoting the readability of analysis codes. Broadly, each class shares a similar usage syntax. Users first initialize an object for the desired algorithm and next provide required parameters. Here, required parameters typically include a state vector ensemble ( $X_p$ ), proxy records ( $Y$ ), proxy estimates ( $\hat{Y}$ ) or forward models, and proxy error-variances or covariances ( $R$ ). Users can specify any additional parameters, and then implement the algorithm using the

run method. To support the use of large state vector ensembles, all three algorithms are optimized for both speed and efficient use of memory.

The `kalmanFilter` class is expected to be the most frequently used DA algorithm in the DASH toolbox and contains options for both offline and online regimes. The class implements an ensemble square-root Kalman filter (Andrews, 1968), which processes ensemble means and deviations separately. This separation precludes the need for perturbed observations (Whitaker and Hamill, 2002) and provides several opportunities for enhanced computational efficiency. For example, exploratory analyses can choose to only assimilate the ensemble mean, which is significantly faster than updating the full ensemble. Other optimizations leverage the independence of deviation updates from the proxy records to minimize the number of computations of the Kalman Gain. The `kalmanFilter` class also supports many options for adjusting Kalman filter covariance matrices (the  $\text{cov}(X, \hat{Y})$  term in Equation C.5); these include covariance inflation (Anderson and Anderson, 1999), localization (Hamill et al., 2001), and blending ensemble covariances with a second covariance matrix (e.g. Valler et al., 2019). The class also permits user-specified covariance matrices, which can be useful for changing continental configurations in deep-time assimilations. Finally, the class supports the use of evolving offline priors (e.g. Franke et al., 2020; Osman et al., 2021), which can be used simulate changing climate system boundary conditions while minimizing computational cost.

Naïve Kalman filters return an entire state vector ensemble in each assimilated time step, which can rapidly exceed computer memory. Consequently, the `kalmanFilter` class includes many options for reducing the size of the outputs. Alternatives to saving full ensembles include only returning the ensemble mean, returning the ensemble mean and variance, and returning several percentiles of the full ensemble. The class also provides support for reconstructing climate indices from assimilated spatial fields while conserving computer memory. In many cases, an assimilated spatial field is primarily used to calculate a reconstructed climate index. The full posterior of a climate index is often useful for uncertainty analysis, but spatial fields are often too large to allow the return of full posterior ensembles. To remedy this situation, the `index` method allows users to calculate and return the full posterior of a climate index, without saving the full-field

posterior ensemble.

The `particleFilter` class provides an alternative algorithm to Kalman filtering. In DASH, this algorithm proceeds by weighting the state vectors (particles) in an ensemble and then computing a weighted mean across the ensemble. The primary option in the `particleFilter` class concerns the method used to determine the weights for the mean. By default, the class implements a Bayesian weighting scheme that conforms to a classical particle filter (see Van Leeuwen, 2009). However, users can instead choose to take a mean of the best  $N$  particles, with the number of particles determined by the user.

The `optimalSensor` class is based on the method described by Comboul et al. (2015), which is derived from an ensemble Kalman filter framework. Rather than reconstructing climate variables over time, the algorithm instead tests the ability of a proxy record to reduce the variance of a climate metric calculated over an ensemble. Essentially, this method assesses the relative influence of individual proxy records on a reconstructed index. The `optimalSensor` class provides three distinct, yet related, routines to support these types of analyses. The `evaluate` routine allows users to assess each proxy's individual ability to reduce variance in the posterior ensemble. The `run` routine implements the greedy algorithm of Comboul et al. (2015), and allows users to rank the utility of proxy sites for successive assimilation. Finally the `update` routine assesses the total variance reduced by an entire proxy network. These commands can also be combined to examine changes in proxy influence as additional records are added to a network.

Classically, the optimal sensor algorithm strictly requires proxy error-variances, which necessarily assumes that assimilated proxy records are independent. However, the `optimalSensor` class extends the algorithm to allow for covarying proxy errors. In this case, the covarying-proxies are processed using a single block-update, effectively treating the covarying proxies as a single sensor. This is useful when assimilating gridded, spatially-covarying proxy networks and climate field reconstructions (e.g. King et al., 2022), such as drought atlases (e.g. Cook et al., 1999, 2010; Morales et al., 2020).

## C.5 Examples

In this section, we provide two examples illustrating the use of the DASH toolbox. These examples are designed to demonstrate the use of DASH for a variety of analyses over different spatial scales, time periods, and proxy networks. These examples closely mimic several existing studies in the paleoclimate DA literature (King et al., 2021; Tierney et al., 2020b; Osman et al., 2021), although we have modified the analyses at several points for brevity or to demonstrate extended capabilities of the DASH toolbox. Numbers in parentheses refer to the line numbers in the code for each example.

### C.5.1 Northern Hemisphere Summer Temperatures over the Last Millennium

Our first example illustrates one possible setup for reconstructing summer temperatures in the extratropical Northern Hemisphere over the last millennium using annual resolution proxies. This example follows the assimilations found in King et al. (2021), although for the sake of simplicity, we only assimilate a single climate model here. In this example, we integrate a network of 54 temperature-sensitive tree-ring records (Wilson et al., 2016; Anchukaitis et al., 2017) with output from the CESM1.1 Last Millennium Ensemble (Otto-Bliesner et al., 2016) to reconstruct both a summer temperature spatial field and a spatial-mean index. We generate proxy record estimates using simple linear forward models trained on the mean temperature of each site’s optimal growing season. We run the assimilation using an ensemble Kalman Filter with a stationary offline prior. We also apply covariance localization for the spatial field, which we implement using a Gaspari-Cohn 2D polynomial (Gaspari and Cohn, 1999) with a 20,000 km cutoff radius. Finally, we use an optimal sensor analysis to evaluate the potential influence of each tree-ring record in the network. The results of this analysis are displayed (Figure C.3) using the visualization codes in the this paper’s data repository.

#### Organize Climate Data

The first two sections of the example (lines 6-52) illustrate using `gridfile` to organize data used in the assimilation. Here, this data consists of (1) climate model output from

the CESM1.1 Last Millennium Ensemble and (2) tree-ring chronologies. We begin with the climate model output. In this example, the climate model output consists of reference height temperatures from fully-forced run #2. This output is stored across two NetCDF files and spans a 2D spatial grid over the period 850 to 2005 CE at monthly resolution. Our first step is to create a metadata object that defines the scope of this dataset (lines 12-18). Here, we choose to define spatial metadata using the latitude and longitude values stored in the NetCDF output files (lines 13-14). However, the time metadata in the NetCDF files is reported as “days since January 1, 850”, which is non-intuitive for our purposes. Instead, we choose to define time metadata using MATLAB’s built-in `datetime` format, which will allow us to sort time points by months and years (line 15). We also include two optional metadata attributes to better document the dataset (line 18). We next create a `gridfile` object whose scope is defined by this metadata (line 21) and add the temperature dataset, stored in the `TREFHT` variable of the two NetCDF files, to the `gridfile` object’s catalogue (lines 28-29). Finally, we apply a data transformation to the catalogue (line 32) so that loaded temperature data will be returned in units of Celsius, rather than Kelvin.

In the next section (lines 35-52), we catalogue the tree-ring chronologies. These records are stored in a binary MAT-file (line 38), along with information about each proxy site. The proxy record dataset can be described as a 2D array that spans the 54 proxy sites over time at annual resolution, and we first define metadata for this dataset (line 43). Here, we choose to define metadata along the proxy-site dimension using the ID, spatial location, and optimal growing season of each site (line 42). For time metadata, we use the calendar year corresponding to each measurement (line 43). We next create a `gridfile` object whose scope is defined by this metadata (line 46) and add the proxy record dataset, stored in the `crn` variable of the MAT-file, to the `gridfile` catalogue (line 47). Finally, we indicate that -999 values in the dataset represent fill values and should be converted to NaN when loaded (line 50).

### **Build a State Vector Ensemble**

In the next section (line 53-101), we use the `stateVector` class to design and build a state vector ensemble. We begin by initializing and labeling a `stateVector` object (line 56), and then initializing variables within that state vector (lines 62-63). Typically, a state vector will include any variables required to run the proxy system forward models, as well as reconstruction targets. In this example, each proxy system model requires a seasonal temperature mean from the model grid point closest to the proxy site. Thus, we first initialize variables for the temperature means of the proxy records using a different variable name for each site (line 62). We use a different variable name for the variable at each site. We also create variables for the reconstructed spatial temperature field, and the spatial-mean index (line 63), for a total of 56 variables. All of these variables will be constructed from the monthly LME temperature output, which is indicated by the second input in lines 62 and 63. Note that the names of state vector variables do not need to match the names of variables stored in data source files – here, `TREHFT` – because multiple state vector variables may be derived from the same dataset.

We next specify how to select ensemble members in the state vector ensemble. In this example, we indicate that ensemble members should be selected along the time dimension, with each ensemble member associated with a particular calendar year (line 69). Using 0 as the first input applies this setting to every variable in the state vector. Here, we use January as a reference point for each calendar year, but this does not imply that the variables will necessarily contain data from the month of January. Instead, the January months are used to align variables so that the values within any given ensemble member correspond to the same year. For example, consider two variables implementing seasonal means. One variable, `MJJA`, implements a seasonal mean from May to August. The other variable, `ON`, implements a seasonal mean from October to November. Although the two variables cover different seasonal windows, the seasonal windows for each ensemble member should be drawn from the same year. Here the January reference point allows us to ensure that these seasonal windows are aligned to the same year; essentially, the variables for each ensemble member will be built using the first appropriate seasonal

window following the associated January reference point. For an ensemble member that uses January 1850 as a reference point, the MJJA variable will be built using data from May-August 1850, and the ON variable will be built using data from October-November 1850. Although the two variables use different temporal spans, they collectively refer to the same year within the ensemble member. Additionally, the state vector class will ensure that ensemble members are only selected from years that include complete temporal spans for all variables. Continuing the example: if the temperature dataset ended in October 1900, then 1900 will never be selected as an ensemble member, because the ON variable would be missing data from November of that year.

Finally, we design the variables so that each uses values from the appropriate subset of the monthly temperature dataset. For the reconstruction targets, we use grid points from the extratropical Northern Hemisphere (line 78) and summer (June-August) seasonal temperature means (line 79). We note that the third input in line 78 is left empty because the latitude dimension should not be used to select different ensemble members (contrast this with the time dimension in line 69). To implement the seasonal means, we provide the indices of months relative to each January reference point. As the reference point, each January is given a relative index of 0; hence, a June-August mean is calculated using data values 5, 6, and 7 (monthly) time steps after each January reference point. We also specify a latitude-weighted spatial mean for the spatial-mean index (line 80). Before designing the forward-model variables, we first note that each variable uses a different seasonal average. Including the full spatial field for multiple different seasonal windows would result in an unnecessarily large state vector, so we first use the `closestLatLon` DASH utility to locate the model grid point closest to each proxy site (line 86). We then design each forward-model variable to consist of the site-specific seasonal temperature mean at that single grid point (lines 88-96). At this point, we have finished designing the state vector, and proceed to build an ensemble with 1000 members (line 99). In this example, we save the built ensemble to a `.ens` file. Although the `stateVector` class can also return ensemble directly as output, we generally recommend saving to file, because this allows the DASH toolbox to use computer memory more efficiently.

## Proxy Forward Models

The next section (lines 102-123) uses the `PSM` package and `ensembleMetadata` class to design proxy forward models and run the models on values stored in the state vector ensemble. The outputs of these forward models are the proxy estimates used to compare state vector ensemble members to observed proxy records in assimilation algorithms. We begin by using the `PSM` package to create simple, linear forward models for each proxy site (line 112). The coefficients for each model are calibrated to mean temperature over the optimal growing season at each proxy site. Determining forward model coefficients is beyond the scope of this example, but King et al. (2021) compute these values by regressing the proxy records against an instrumental temperature dataset. After designing each model, we next indicate the state vector row that corresponds to the inputs for each model (lines 116-117). Finally, we use the `estimate` command to run the forward models on the ensemble and generate the proxy estimates (line 121).

## Kalman Filter

In this section (lines 124-172), we use the `kalmanFilter` class to implement an ensemble Kalman filter and reconstruct summer temperatures. We first initialize and label a `kalmanFilter` object, which will store the parameters used to run the assimilation. The mandatory parameters for an ensemble Kalman filter are (1) a prior ensemble, (2) proxy records, (3) proxy estimates, and (4) proxy error covariances or variances, and we next provide these parameters to the `kalmanFilter` object (lines 132, 136, 137, and 141). Determining proxy error variances is beyond the scope of this example, but King et al. (2021) compute these values by running the proxy forward models on an instrumental temperature dataset and comparing the resulting proxy estimates to the real proxy records. In this example, we also implement covariance localization. To accomplish this, we first calculate localization weights for the ensemble (line 146) and proxy sites and then provide these weights as parameters to the `kalmanFilter` object (line 147).

To illustrate the flexibility of the DASH architecture, we also demonstrate a second method for reconstructing the spatial-mean summer temperature index (line 154). This

method allows the user to calculate an index from the posterior of a spatial field, without saving the (often very large) spatial field posterior. To further conserve memory, we also indicate that the filter should only record the variance and percentiles of the posterior ensemble (lines 157-158), rather than the much larger full posterior. Finally, we run the Kalman filter algorithm for the analysis and return the mean, variance, and posterior mean and percentiles of the target reconstruction variables (line 161). We note that the reconstructed spatial field is organized as a state vector, but many mapping functions operate on spatial matrices, rather than vectors. Hence, to facilitate display of the reconstructed spatial field, we regrid the field posterior to the spatial dimensions of the original climate model output (lines 165-166). We also extract the assimilated spatial temperature mean, which is the final element along the state vector (line 167), and alternate spatial mean calculated from the updated spatial field (line 168). Figure C.3 illustrates the results of this assimilation. The upper panel compares the reconstructed indices obtained using the two different methodologies: the blue line depicts the index obtained by assimilating the temperature spatial mean directly in the state vector, and the red line depicts the index calculated from the updated (posterior) spatial field. The lower left and lower center panels display an example reconstructed spatial field from 1850 CE along with an uncertainty quantification based on the variance of the field's posterior ensemble. Notably, the spatial indices calculated using the two different methods are not identical. In brief, this discrepancy occurs because (1) the index calculated from the posterior field (in red) is sensitive to spatial heterogeneity in the Kalman filter updates, and (2) the directly assimilated index (in blue) is less sensitive to the proxy records than are individual spatial sites. The causes and implications of this behavior are discussed in greater detail in Section C.6.3.

### **Optimal Sensor**

In the final section (lines 174-195), we use an optimal sensor framework to evaluate the influence of each proxy on the reconstructed spatial-mean index. Analogous to the `kalmanFilter` object of the previous section, here we will use an `optimalSensor` object to organize parameters for the analysis. The required parameters for an optimal sensor are (1) a prior ensemble, (2) proxy estimates or forward models, (3) proxy error-

variances, and (4) a sensor metric. After initializing and labeling the `optimalSensor` object (line 177), we set the extratropical summer temperature field as the prior (lines 181-182) and also provide proxy estimates and error-variances (lines 183). We next specify the latitude-weighted spatial-mean of the summer temperature field as the sensor metric (lines 187-188). With these parameters set, we then use the optimal sensor to evaluate the power of each proxy for reconstructing the spatial-mean index (lines 191). Figure C.3 (lower right) displays the results of this analysis. Here, the ability of a proxy to reduce variance responds to two factors: the covariance of its estimates with the modeled spatial-mean index, and its uncertainty values ( $R$ ), which represent the accuracy of its forward model. Thus, the proxies with the greatest ability to reduce variance are characterized by more accurate forward models and stronger covariance with the spatial-mean index.

### C.5.2 Global Sea Level Pressures at the Last Glacial Maximum

Our second example illustrates a setup for reconstructing global sea level pressures from the Last Glacial Maximum (LGM) to present at 3,000-year resolution. This example is inspired by Osman et al. (2021) with several modifications. First, we assimilate global sea level pressures rather than sea surface temperatures (SSTs) in order to demonstrate the reconstruction of climate variables not directly sensed by the proxy network. For the sake of simplicity, we also limit the proxy network to the alkenone  $U_{37}^{K'}$  and  $\delta^{18}O$  of planktic foraminifera SST proxies, neglect spatial variations in proxy seasonal sensitivities, and reconstruct spatial fields on a 3,000-year time step. In this example we integrate a network of  $U_{37}^{K'}$  and  $\delta^{18}O$  sediment records with output from the isotope-enabled Community Earth System Model (iCESM1.2; Brady et al., 2019; Tierney et al., 2020b; Zhu et al., 2017; Stevenson et al., 2019). We generate proxy record estimates using the `BayFOX` (Malevich et al., 2019) and `BaySPLINE` (Tierney and Tingley, 2018) forward models. We conduct the assimilation using an ensemble Kalman filter with an evolving offline prior and also implement a proxy-validation analysis. The results of this analysis are displayed in Figure C.4 using the visualization codes in this paper's data repository.

## Organize Climate Data

Similar to Example 1, the first two sections again use the `gridfile` package to organize climate data. Here, the data consists of (1) climate model output from iCESM binned to 50-year intervals, and (2)  $U_{37}^{K'}$  and  $\delta^{18}\text{O}$  proxy records. The climate model output includes variables for the sea-level pressure (SLP) reconstruction target, as well as sea surface temperatures (SSTs) and  $\delta^{18}\text{O}_{sw}$ , which are used to run the proxy forward models. The variables are binned to monthly 50-year averages in order to more closely match the multi-decadal averages captured by the proxy records. This output includes sixteen 50-year averages for each 3,000-year interval from the LGM to present. The data for the variables is stored in three separate NetCDF files. The SLP variable is provided on the model's rectilinear atmosphere grid, and the creation of its `gridfile` catalogue (lines 10-14, 23, 28) follows the process outlined in section 4.1.1. In contrast, the SST and  $\delta^{18}\text{O}_{sw}$  variables are sourced from the ocean component of the model, which uses a tripolar coordinate system. Tripolar datasets typically include dimensions for both latitude and longitude, but spatial metadata is not fixed for any given element of either dimension. For example, the latitude value at  $(\text{latitude}_j, \text{longitude}_k)$  is not the same as the latitude value for  $(\text{latitude}_j, \text{longitude}_{k+1})$ . Consequently, the dataset describes values at distinct (latitude, longitude) points, rather than values on a rectilinear (latitude x longitude) grid. The `gridfile` class requires fixed metadata values along each data dimension, so we define the metadata for SST and  $\delta^{18}\text{O}_{sw}$  using unique spatial sites (lines 16-20), rather than a rectilinear latitude x longitude format. Note on lines 29 and 30 that two dataset dimensions are associated with the site spatial dimension. This syntax merges the latitude and longitude dimensions in the `gridfile` catalogue and treats them as a single spatial dimension. We next use `gridfile` to catalogue the proxy records (lines 33-48). Here, the proxy records are stored in a delimited text file, and the first line of the file reports proxy record IDs, rather than data values. Metadata for the proxies is stored in a separate text file, and we use this metadata to indicate the ID, spatial coordinates, proxy type ( $U_{37}^{K'}$  or  $\delta^{18}\text{O}$ ), and foraminiferal species associated with each record (line 42).

## State Vector Ensemble and Evolving Prior

We next design and build the state vector ensemble for the LGM assimilation (lines 49-83). We begin by initializing a `stateVector` object with three variables (lines 52 and 60). The first variable, SLP, is the reconstruction target; the other two variables, SST and  $\delta^{18}\text{O}_{sw}$ , are required to run the proxy forward models. We next indicate that ensemble members should be selected from different points in time with each ensemble member associated with a particular binned 50-year interval (line 65); here, we use January bins as a reference point for each 50-year interval. We emphasize that the climate model datasets are *monthly* 50-year averages, and using January as a reference point allows us to select different monthly bins from within the same 50-year interval. In this example, we target annual SLP values, so we use an annual mean of the 50-year bins for SLP (line 70). Similarly, we use an annual mean of the  $\delta^{18}\text{O}_{sw}$  output (line 70), because the proxy forward models require annual  $\delta^{18}\text{O}_{sw}$  values. By contrast, the proxy models require monthly SST inputs, so we use the `sequence` command to include data from each monthly 50-year SST average in the state vector (line 71).

We note that, unlike Example 1, we do not design variables for the individual proxy records; instead, we include the entire spatial field for each climate variable used by the forward models. This syntax simplifies the code but results in a larger state vector. We elect to use this syntax here in order to improve code clarity and also demonstrate the flexibility of the DASH architecture. However, other applications should compare the benefits of code clarity with greater memory use when choosing a syntax. Finally, we build a state vector ensemble using all available ensemble members (line 74). We select ensemble members sequentially in order to facilitate the creation of an evolving prior. This orders the ensemble members so that the 50-year averages for each 3,000-year interval are all in succession. We next use the `evolving` command to implement an evolving prior for the different 3,000-year intervals (line 81). For this command, the columns of the `members` variable indicate which ensemble members should be used for each evolving prior. Here, each prior is built using the 16 50-year bins for one of the nine 3,000-year intervals.

## Proxy Forward Models

We next build and run proxy forward models on the state vector ensemble in order to generate a set of proxy estimates. Here, we use the `BaySPLINE` and `BayFOX` Bayesian forward models for  $U_{37}^{K'}$  and  $\delta^{18}\text{O}$ , respectively. We begin by using the `download` command to download the models from their respective Github repositories and add them to the MATLAB active path (lines 87-88). We next design a forward model for each proxy record using the model appropriate for each proxy's type (lines 90-120). For the `BaySPLINE` model, we locate state vector rows corresponding to the monthly SST values from the climate model grid point closest to each proxy record (lines 102-104). The `BayFOX` model is calibrated to different foraminiferal species, so we initialize each model with the species of the associated proxy record (line 109). We then locate monthly SST values and annual-mean  $\delta^{18}\text{O}_{\text{sw}}$  values, again at the closest climate model grid point (lines 112-114). For the purposes of documentation, we also label each forward model with the ID of the associated proxy record (line 118). Finally, we run the forward models on the evolving state vector ensemble using the `estimate` command (line 124). In addition to proxy estimates, the `BaySPLINE` and `BayFOX` models calculate proxy error-variances, which are provided as the second output.

## Kalman Filter and Proxy Validation

We next implement the Kalman Filter analysis (lines 127-152). We first initialize and label a `kalmanFilter` object (lines 130) and then provide the required algorithm parameters (lines 135-138). To conserve memory, we only return the mean and variance of the posterior ensemble (line 142). As in Example 1, we regrid the reconstructed spatial field to the dimensions of the original climate model to support visualization and post-processing (lines 146-147; Figure C.4). Unlike Example 1, we include all of the climate variables needed for the proxy forward models in the prior. This allows us to run the proxy forward models on the reconstruction and generate proxy posterior estimates. We can then compare these estimates to the real proxy records as a basic assessment of reconstruction skill (Figure C.4). We implement this process by applying the `estimate`

command to the posterior (line 156). For the sake of brevity, we only implement a simplified proxy validation in this example. In practice, DA applications should validate the reconstruction using proxies withheld from the assimilation (e.g. Tierney et al., 2020b; Osman et al., 2021; King et al., 2021) so that assimilated proxies do not inform the skill of their own validation values.

### **C.5.3 Additional Considerations**

The examples presented above touch upon many aspects of paleoclimate DA workflows but cannot be exhaustive. For the sake of brevity and clarity we have neglected several considerations common in DA applications. One particular step we have neglected is the determination of proxy uncertainties ( $R$  in equations C.3, C.5, and C.6). In some cases, proxy uncertainties ( $R$ ) may be provided by the proxy forward models (as in Example 2) or from the calibration of the forward models (e.g. Tardif et al., 2019; King et al., 2021). Another potential approach involves running the forward models on instrumental data and comparing the resulting proxy estimates to the real proxy records (e.g. King et al., 2021, 2022). However, we note that these approaches are not applicable to all analyses, so users may need to develop additional methods to estimate proxy uncertainties. For example, methods that estimate proxy error-variances (e.g. Tardif et al., 2019; Tierney et al., 2020b; King et al., 2021) implicitly assume the independence of proxy uncertainties. However, this assumption may not hold when proxy records are strongly correlated or sensitive to the same local climate variables; when this occurs, proxy error-covariances should be used in place of error-variances (see King et al., 2022, for an example). We also discuss additional issues common to many paleoclimate applications in the section below.

### **C.6 Warnings and Best Practices**

While it is not possible to detail all the issues that can occur when using DA for paleoclimate reconstructions, here we mention several cautions and suggestions for best practices. Along with methodological considerations, DA users should be aware of the limitations

of both the proxy data and prior modeled climate states. In other words, simply running an assimilation code does not guarantee that a reconstruction is scientifically valid, and potential DA users should understand the tradeoffs and issues with DA methods when designing a reconstruction. In this section, we present several major challenges that may be encountered in paleoclimate DA, and we outline approaches to mitigate or recognize their effects. This list is by no means exhaustive, and we strongly recommend that potential DASH users first familiarize themselves with the paleoclimate DA literature and also evaluate their reconstructions for sensitivity to the assumptions and input data.

### **C.6.1 Temporal Variability**

A major issue when using an ensemble Kalman filter with a static prior (e.g. Steiger et al., 2014; Hakim et al., 2016; Dee et al., 2016; Tardif et al., 2019; Steiger et al., 2018; Neukom et al., 2019; Zhu et al., 2021a; King et al., 2021, 2022) is that the proxy network’s size and composition – and changes to these properties over time – can directly alter the temporal variability of the reconstruction. Essentially, we have observed that variability is artificially reduced as the proxy network becomes smaller. For a changing sample size, a common feature of paleoclimate reconstructions, the reconstruction’s variability is thus non-stationary and relative climate variability may not remain consistent over the span of the reconstruction. This is a major concern for paleoclimate studies that seek to use reconstructions to identify and interpret evidence of changing climate variability through time.

This effect occurs because a static prior implies zero temporal variability as an *a priori* assumption in the absence of proxy information. Consider a “no-information” case, in which a static prior is assimilated with an empty proxy network. Since the proxy network is empty, the prior ensemble will not receive any updates, and the reconstruction will be the mean of the prior in every time step. Since the prior is identical in every time step, the reconstruction will consist of a constant value over time and will exhibit no temporal variability. With the addition of a proxy record to the network, the prior will begin to receive updates, and the reconstruction will begin to gain temporal variability. Each subsequent record added to the proxy network increases the ability of the method to

move the reconstruction off the prior mean, and so reconstruction variability will increase with the size of the proxy network. This behavior is by design: in the absence of the additional information, the prior provides the best estimate of the mean state of the climate system. However, it creates complications for paleoclimate interpretations. We note that this effect is most severe for smaller proxy networks and at spatial points informed by a limited number of proxy records.

Because of this effect, it is essential that assimilations using static priors account for the effects of proxy network composition on temporal variability. Variance adjustment methods are common in other approaches to paleoclimate reconstruction (e.g. Cook et al., 1999; Esper et al., 2005; Frank et al., 2007; Anchukaitis et al., 2017), and King et al. (2022) provide an example for how this can be accomplished for DA applications. Alternatively, evolving priors can mitigate the variance issue (e.g. Tierney et al., 2020b; Osman et al., 2021) by removing the *a priori* assumption of zero temporal variability. However, we caution that even evolving priors can exhibit a variance dampening effect when the variability between reconstruction time steps and the state of the evolving priors is dominated by internal climate variability.

### **C.6.2 Climate Model Biases**

A second major concern for paleoclimate DA concerns the effects of climate model biases on assimilated reconstructions. In this discussion, we find it useful to distinguish between (1) biases in the mean state, and (2) climate model covariance biases. Mean state bias refers to the systematic tendency of a simulated variable to be too high or too low compared to observations. Covariance bias refers to errors in the linear relationship between climate variables at different spatial points, or between different variables. Essentially, these are biases in the teleconnection patterns associated with various climate phenomena. Since the model prior covariance determines how information propagates from a proxy network to distal parts of a climate field, differences between the real and modeled climate system covariance will cause errors in the assimilation. No climate model can match the complexity of the real Earth system and so all climate models necessarily include some of these errors.

An additional consequence of climate model biases concerns method testing and proof-of-concept studies for paleoclimate DA. Typically, these studies rely on pseudo-proxy frameworks, in which climate model output is used to simulate a set of proxy records. These pseudo-proxy records are designed to mimic a real proxy network and can be used to reconstruct the climate model output. Unlike the real past climate history, the climate model output is fully known and so provides an opportunity to assess assimilation skill. Due to the complexity of these studies, it can be tempting to use the same climate model to both generate the pseudo-proxies and build the assimilation prior. However, we caution that this framework represents an unrealistic “perfect-model” design, in which the climate model used for assimilation perfectly describes the target climate system. Although perfect-model experiments have their uses, climate model biases represent a major source of error in paleoclimate DA (Dee et al., 2016; King et al., 2021) and DA users should account for these biases to accurately quantify DA skill. Ultimately, “biased-model” experiments, which use different climate models to generate pseudo-proxies and build the assimilation prior, are necessary for accurate method testing. We also note that the exact nature of climate model biases will vary by model and the specific climate variable, and so an ensemble of different biased-model tests is often necessary to capture the full effects of climate model biases.

Deleterious effects in real assimilations also occur when the inputs to the proxy forward models exhibit mean state biases. These biases propagate through the forward models to the proxy estimates and thereby influence the comparison of the ensemble members to the real proxy records. In some cases, this can cause artificial trends in a reconstruction. Essentially, the assimilation draws reconstructed variables unilaterally in the direction of less biased mean values. Although this does indeed improve the final estimate of a variable’s value, this behavior is mixed with the variable’s reconstructed temporal evolution and causes an artificial trend. Simultaneously, mean state biases may cause severe errors in proxy forward models. This issue is typically most extreme for forward models that rely on calibrations or process thresholds based on absolute units. For example, the VS-Lite tree-ring model (Tolwinski-Ward et al., 2011) includes a temperature threshold based on absolute Celsius units. At temperatures below this threshold, VS-Lite assumes

no growth occurs and produces a proxy estimate of zero. As a result, climate models with a cold bias may consistently fall below this threshold, causing VS-Lite to produce a null record. In this case, as a consequence of the mean state bias, VS-Lite would assume that trees cannot grow at a location where they do grow in reality, and this error would degrade the reconstruction.

Some mean state biases can be addressed by the process of bias correction used in other disciplines and applications (e.g. Wang and Robertson (2011); Zhao et al. (2017); Cannon et al. (2015); Cannon (2018); Galmarini et al. (2019), and see Steiger et al. (2018) for a DA example). When appropriate, users can alternatively avoid the effects of mean state biases by providing climate anomalies to the proxy forward models, rather than absolute values (e.g. Tardif et al., 2019; King et al., 2021, 2022). This is often appropriate for assimilations that rely on linear proxy forward models or forward models not dependent on absolute units. If using priors from multiple climate models, users may also need to avoid or account for time periods when climate models strongly differ, as strongly differing climate representations can act analogously to mean state biases. For example, the instrumental era is often not suitable for computing climate model anomalies for long preindustrial and last millennium simulations, because the climate response to anthropogenic influences can vary widely across models. By contrast, anomalies assessed relative to the pre-industrial period are typically more stable.

Covariance biases are perhaps the more challenging to deal with since they bias the propagation of information from the proxy records to the reconstruction targets and do not present simple fixes. Multivariate bias correction methods may provide a solution to this issue (e.g. Cannon, 2018; Vrac, 2018; Galmarini et al., 2019), but these methods have seen little use in paleoclimate DA contexts. Instead, a more common solution is to assimilate a multi-model ensemble (Parsons et al., 2021; King et al., 2021, 2022). Users may enact this using a single multi-model prior (e.g. Parsons et al., 2021; King et al., 2022), or by performing an ensemble of assimilations using different single-model priors (e.g. King et al., 2021). When possible, we recommend the use of multi-model priors. These priors are supported in the DASH framework, and they limit the effects of covariance biases by down-weighting covariance patterns that disagree across different models. We also note

that this down-weighting may in part contribute to spatial heterogeneity in Kalman filter updates, which we discuss in detail in the next section.

### **C.6.3 Physically Inconsistent Reconstructions**

Both the particle filter and Kalman filter frameworks assume that all state vector variables and proxy estimates follow a Gaussian distribution; however, not all climate variables meet this criteria. Thus, DA users should take care to transform non-Gaussian variables into an approximately Gaussian space before assimilation. Failing to take this step can result in unrealistic or nonphysical reconstructed values. This is often relevant when assimilating variables distributed near the lower bounds of their domains. For example, precipitation variables often have high probability near zero, yet cannot fall below the lower bound of zero itself, and this results in a strongly non-Gaussian shape. Because of this, raw precipitation values are not suitable for assimilation and using them can cause the method to return negative precipitation values. Thus, users should transform precipitation into an approximately Gaussian shape before assimilation. The reverse transformation can then be applied to the assimilated variables in order to obtain reconstructed precipitation. Transforms such as the extended Box-Cox can be useful for adjusting variables near a lower bound, but the most appropriate transforms will vary by application (Wang et al., 2012).

We also emphasize that the DA algorithms described in this paper do not conserve physical properties like mass or energy. Consequently, assimilated reconstructions are not bound by the governing equations inherent to the climate models used to generate a prior ensemble and can produce physically inconsistent values. In some cases this may mean that assimilated fields are not suitable for providing boundary conditions for climate model simulations. Unrealistic values can also arise when individual proxy records are given excessive weight in the Kalman filter. When the magnitudes of proxy weights are too large, small proxy innovations can result in drastically large updates to assimilated climate variables. This issue most commonly occurs when proxy uncertainties ( $R$ ) are severely underestimated. For example, in Example 1 our proxy uncertainties incorporate both forward-model errors and non-climatic noise in the proxy records. However, if

we neglect these effects and compute  $R$  using only the uncertainties inherent in measuring tree-ring variables (which are vanishingly small), the resulting Kalman filter updates alter the assimilated temperature field by thousands of degrees Kelvin, a clearly unrealistic result. This behavior underscores the importance of correctly incorporating multiple sources of error when quantifying proxy uncertainties. Although DA methods that conserve physical properties do exist, these methods have seen little use in paleoclimate contexts, likely due to the prevalence of offline regimes.

A related issue concerns the spatial heterogeneity of Kalman filter updates, which can also result in physically inconsistent behavior. When assimilating spatial climate fields, the magnitudes of Kalman filter updates often vary unevenly across different spatial points. The magnitude of the update at a given spatial point is proportional to that point's covariance with the proxy estimates, so distant spatial points that covary less strongly with the proxy network will receive smaller updates. As a result, reconstructed values at distant sites tend remain closer to the prior ensemble mean and exhibit lower temporal variability than sites closer to the proxy network. This lower variability is not a real climate phenomenon, but rather a consequence of the Kalman filter method, which is designed to estimate mean states rather than temporal variability. However, we also note that variance of the posterior ensemble is available for the user to assess the uncertainty resulting from smaller updates.

This spatial heterogeneity also has consequences for reconstructing large-scale climate indices, such as those used to characterize large-scale climate modes and spatial averages. These large-scale indices are typically computed using values from multiple points in a spatial climate field; however, the uneven application of Kalman Filter updates to different spatial points can skew the calculation of these indices. For example, consider the Southern Annular Mode (SAM): a commonly used index to measure the SAM's phase is defined as the gradient of zonal mean sea level pressures between  $40^{\circ}\text{S}$  and  $65^{\circ}\text{S}$  (Gong and Wang, 1999). Now consider an assimilation that uses a proxy network primarily located near  $65^{\circ}\text{S}$ . Because of the proxy network's location, spatial points near  $65^{\circ}\text{S}$  will receive larger updates than those near  $40^{\circ}\text{S}$ ; by contrast, points near  $40^{\circ}\text{S}$  will be less altered and will remain close to the mean of the prior. As a consequence of this effect,

a SAM index determined from the posterior spatial field using this network might only reflect changes to values at  $65^{\circ}\text{S}$ , thereby failing to assess changes at the northern end of the gradient. Thus, when reconstructing climate indices from posterior spatial fields, it is essential for DA users to demonstrate the homogeneity of update magnitudes at the spatial points used to calculate the index. An alternative approach to reconstructing climate indices is to include the climate index directly in the state vector, which precludes the issue of spatial heterogeneity. A tradeoff of this approach is that proxy records will covary less strongly with large-scale indices than with local climate variables, and so reconstruction uncertainty may remain higher overall. However, in the case of spatial heterogeneity, we emphasize that higher uncertainties are preferable to a physically implausible reconstruction.

### **C.7 Discussion**

In this paper, we have presented DASH, a MATLAB toolbox facilitating paleoclimate data assimilation. Unlike existing paleoclimate DA codes (e.g. LMR (Hakim et al., 2016; Tardif et al., 2019), PHYDA (Steiger et al., 2018)), DASH is not built around any one particular analysis. Instead, the toolbox is generalized for a wide variety of analyses and supports DA for a variety of time scales, spatial regions, proxy networks, data formats, and algorithms. The code is highly modular and provides users with fine control over the intricacies of any particular assimilation. The modular nature also allows users to incorporate novel methods not currently supported by the DASH toolbox into a given workflow. For example, proxy system modeling is a field of active research, and users may wish to leverage models not currently implemented in DASH. Rather than limiting users to a strict set of supported models, DASH allows researchers to implement their own proxy estimates outside of DASH entirely. This obviates the need to directly modify the toolbox when developing novel scientific methods. DASH's modular nature also simplifies and clarifies assimilation codes, thereby improving the transparency and reproducibility of associated data assimilation analyses. Additionally, the code allows users to manipulate large datasets using human-readable metadata, which further promotes code

readability. The algorithms in DASH are optimized for both speed and efficient memory use, critical qualities for analyses utilizing multiple spatial fields from high-resolution climate models. Although a general familiarity with the paleoclimate and data assimilation literature is recommended, users do not require previous experience with DA analyses to use DASH.

Because of its flexibility, earlier versions of the DASH toolbox have already been used to implement several paleoclimate reconstructions, ranging across a variety of time scales and reconstruction targets. Tierney et al. (2020b) used a DASH prototype to reconstruct global temperatures at the Last Glacial Maximum using a large proxy network of geochemical SST proxies and model output from iCESM1.2. King et al. (2021) used the toolbox to reconstruct summer temperatures in the extratropical Northern Hemisphere over the last millennium by integrating a temperature-sensitive tree ring network with an ensemble of climate model simulations. Osman et al. (2021) used DASH to produce a full-field reconstruction of surface temperatures from Last Glacial Maximum to present. Rather than conducting a field reconstruction, King et al. (2022) targeted a climate mode index, and reconstructed the Southern Annular Mode over the Common Era using a southern hemisphere proxy network, drought atlases, and a multi-model ensemble. In all four of these studies, DASH was used to implement the assimilation workflow.

### **C.7.1 Future Development**

DASH is an active project and we anticipate continued developments to the toolbox. We welcome suggestions and contributions to the code base. Users interested in contributing are encouraged to either contact the toolbox's developers or submit a pull request to the project's Github repository. We emphasize three major projects for future development here.

First, we note that proxy system modeling is an area of active research. We anticipate the development of new proxy models and recognize the need to incorporate these future models into the DASH framework. The PSM package is built with such development in mind and its modular nature facilitates the incorporation of new models into DASH. Furthermore, DASH includes templates for proxy forward models, thereby allowing users to

incorporate new models into the toolbox as the need arises. Secondly, we note our intention to expand DASH's support of online assimilation algorithms. DASH has primarily been used to implement offline assimilation regimes, and this has influenced the development of the toolbox. Although DASH includes a scaffold for online assimilation, this is limited to a Kalman filter framework and is less optimized than the provided offline algorithms. Avenues for future development include adding frameworks for climate models of varying complexity or emulators and expanding the toolbox's algorithms to include an online particle filter. Finally, we recognize that DASH's reliance on MATLAB precludes a fully open-source toolbox. Although the source code to all DASH analyses is public, the toolbox will not be accessible to users lacking a MATLAB license. Consequently, a final focus of future development involves porting the toolbox to a native Python and/or Julia package.

## **C.8 Conclusion**

In this paper, we describe the features and foundations of DASH, a MATLAB toolbox supporting paleoclimate data assimilation. The toolbox is designed for scripting and command-line use, and helps implement common tasks in paleoclimate data assimilation workflows. Broadly, these include integrating data stored in different formats, designing state vector ensembles, running proxy system forward models, and implementing computationally-efficient data assimilation algorithms. The toolbox provides an interface for external, proxy-system models commonly used in the paleoclimate literature. Data assimilation algorithms in the toolbox include ensemble Kalman filters (both offline and online regimes), particle filters, and optimal sensor analyses. The package is highly flexible and is designed for general paleoclimate data assimilation, rather than any particular DA analysis. As a result of this flexibility, DASH has already been used to implement published paleoclimate reconstructions for a variety of time scales, spatial regions, and proxy networks.

### **Code and Data Availability**

Releases of the DASH toolbox are available on DASH's Github repository (<https://github.com/JonKing93/DASH/releases>) and on MATLAB FileExchange ([reserve-url-pending-review](https://www.mathworks.com/matlabcentral/fileexchange/?term=authorid%3A1507444)). The DASH source code is also available on the Github repository (<https://github.com/JonKing93/DASH>). The input data sets, DASH 4.0.0 release, and visualization codes used in the examples are available in the paper's data repository and at a public Zenodo repository ([reserve-url-pending-review](https://zenodo.org/record/1000000)).

### **Author Contributions**

J.K. wrote the source code for the DASH toolbox. Program conceptualization, the feature set, and the overarching research goals were developed by J.K., K.J.A. and J.E.T. J.E.T., M.O., and K.J.A. tested the toolbox, suggested additional features and improvements, and provided feedback on implementation, documentation, and tutorials. J.K. and M.O. designed the usage examples. All authors wrote the paper.

### **Competing Interests**

The authors declare no competing interests.

### **Acknowledgements**

We thank Jessica Badgeley, Emily Judd, Steven Malevich, and Feng Zhu for fruitful discussions about paleoclimate data assimilation, testing early versions of DASH, and/or suggesting features. We also thank Dave Meko, Sylvia Dee, and Suz Tolwinski-Ward for developing and publishing proxy forward model codes used by DASH. We thank the NCAR modeling group for producing and publishing climate model output used in our examples. The development of DASH was supported by grants from the US National Science Foundation (AGS-1803946, to K.J.A.), grant 2016-015 from the Heising-Simons Foundation (to J.E.T.), and the David and Lucile Packard Foundation (to J.E.T.).

## C.9 References

- Acevedo, W., Fallah, B., Reich, S., and Cubasch, U.: Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model, *Climate of the Past*, 13, 545–557, 2017.
- Alley, R. B.: Palaeoclimatic insights into future climate challenges, *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361, 1831–1849, 2003.
- Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D’Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B. E., et al.: Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions, *Quaternary Science Reviews*, 163, 1–22, 2017.
- Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Monthly weather review*, 127, 2741–2758, 1999.
- Andrews, A.: A square root formulation of the Kalman covariance equations., *AIAA Journal*, 6, 1165–1166, 1968.
- Ault, T., Deser, C., Newman, M., and Emile-Geay, J.: Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium, *Geophysical Research Letters*, 40, 3450–3456, 2013.
- Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., and Meko, D. M.: Assessing the risk of persistent drought using climate model simulations and paleoclimate data, *Journal of Climate*, 27, 7529–7549, 2014.
- Bell, A. R., Osgood, D. E., Cook, B. I., Anchukaitis, K. J., McCarney, G. R., Greene, A. M., Buckley, B. M., and Cook, E. R.: Paleoclimate histories improve access and sustainability in index insurance programs, *Global environmental change*, 23, 774–781, 2013.

- Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, *Climate of the Past*, 8, 963–976, 2012.
- Bradley, R. S.: Are there optimum sites for global paleotemperature reconstruction?, in: *Climatic variations and forcing mechanisms of the last 2000 years*, pp. 603–624, Springer, 1996.
- Brady, E., Stevenson, S., Bailey, D., Liu, Z., Noone, D., Nusbaumer, J., Otto-Bliesner, B., Tabor, C., Tomas, R., Wong, T., et al.: The connected isotopic water cycle in the Community Earth System Model version 1, *Journal of Advances in Modeling Earth Systems*, 11, 2547–2566, 2019.
- Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., and Otto-Bliesner, B. L.: Pliocene and Eocene provide best analogs for near-future climates, *Proceedings of the National Academy of Sciences*, 115, 13 288–13 293, 2018.
- Cane, M. A., Braconnot, P., Clement, A., Gildor, H., Joussaume, S., Kageyama, M., Khodri, M., Paillard, D., Tett, S., and Zorita, E.: Progress in paleoclimate modeling, *Journal of Climate*, 19, 5031–5057, 2006.
- Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, *Climate dynamics*, 50, 31–49, 2018.
- Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?, *Journal of Climate*, 28, 6938–6959, 2015.
- Chen, Z. et al.: Bayesian filtering: From Kalman filters to particle filters, and beyond, *Statistics*, 182, 1–69, 2003.
- Coats, S., Smerdon, J., Stevenson, S., Fasullo, J., Otto-Bliesner, B., and Ault, T.: Paleoclimate constraints on the spatiotemporal character of past and future droughts, *Journal of Climate*, 33, 9883–9903, 2020.

- Comboul, M., Emile-Geay, J., Hakim, G. J., and Evans, M. N.: Paleoclimate sampling as a sensor placement problem, *Journal of Climate*, 28, 7717–7740, 2015.
- Cook, B. I., Cook, E. R., Anchukaitis, K. J., Seager, R., and Miller, R. L.: Forced and unforced variability of twentieth century North American droughts and pluvials, *Climate dynamics*, 37, 1097–1110, 2011.
- Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K.: Drought reconstructions for the continental United States, *Journal of Climate*, 12, 1145–1162, 1999.
- Cook, E. R., Anchukaitis, K. J., Buckley, B. M., D'Arrigo, R. D., Jacoby, G. C., and Wright, W. E.: Asian monsoon failure and megadrought during the last millennium, *Science*, 328, 486–489, 2010.
- Crowley, T. J.: Utilization of paleoclimate results to validate projections of a future greenhouse warming, in: *Developments in atmospheric science*, vol. 19, pp. 35–45, Elsevier, 1991.
- Dee, S., Emile-Geay, J., Evans, M., Allam, A., Steig, E., and Thompson, D.: PRYSM: An open-source framework for PProX System Modeling, with applications to oxygen-isotope systems, *Journal of Advances in Modeling Earth Systems*, 7, 1220–1247, 2015.
- Dee, S. G., Steiger, N. J., Emile-Geay, J., and Hakim, G. J.: On the utility of proxy system models for estimating climate states over the Common Era, *Journal of Advances in Modeling Earth Systems*, 8, 1164–1179, 2016.
- Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, *Climate dynamics*, 38, 527–546, 2012.
- Dubinkina, S. and Goosse, H.: An assessment of particle filtering methods and nudging for climate state reconstructions, *Climate of the Past*, 9, 1141–1152, 2013.
- Esper, J., Frank, D. C., Wilson, R. J., and Briffa, K. R.: Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophysical Research Letters*, 32, 2005.

- Evans, M. N., Kaplan, A., and Cane, M. A.: Optimal sites for coral-based reconstruction of global sea surface temperature, *Paleoceanography*, 13, 502–516, 1998.
- Evans, M. N., Kaplan, A., Cane, M. A., and Villalba, R.: Globality and Optimality in Climate Field Reconstructions from Proxy Data, in: *Interhemispheric Climate Linkages*, edited by Markgraf, V., pp. 53–72, Cambridge University Press, Cambridge, UK, 2001.
- Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Science Reviews*, 76, 16–28, 2013.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, 99, 10 143–10 162, 1994.
- Fang, S.-w., Khodri, M., Timmreck, C., Zanchettin, D., and Jungclaus, J.: Disentangling internal and external contributions to Atlantic multidecadal variability over the past millennium, *Geophysical Research Letters*, 48, e2021GL095 990, 2021.
- Fernández-Donado, L., González-Rouco, J., Raible, C., Ammann, C., Barriopedro, D., García-Bustamante, E., Jungclaus, J. H., Lorenz, S., Luterbacher, J., Phipps, S. J., et al.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, *Climate of the Past*, 9, 393–421, 2013.
- Frank, D., Esper, J., and Cook, E. R.: Adjustment for proxy number and coherence in a large-scale temperature reconstruction, *Geophysical Research Letters*, 34, 2007.
- Franke, J., Valler, V., Brönnimann, S., Neukom, R., and Jaume-Santero, F.: The importance of input data quality and quantity in climate field reconstructions – results from the assimilation of various tree-ring collections, *Climate of the Past*, 16, 1061–1074, <https://doi.org/10.5194/cp-16-1061-2020>, 2020.
- Galmarini, S., Cannon, A. J., Ceglar, A., Christensen, O. B., de Noblet-Ducoudré, N., Dentener, F., Doblas-Reyes, F. J., Dosio, A., Gutierrez, J. M., Iturbide, M., et al.: Ad-

- justing climate model bias for agricultural impact assessment: How to cut the mustard, *Climate services*, 13, 65–69, 2019.
- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757, 1999.
- Gong, D. and Wang, S.: Definition of Antarctic oscillation index, *Geophysical Research Letters*, 26, 459–462, 1999.
- Goosse, H., Renssen, H., Timmermann, A., Bradley, R. S., and Mann, M. E.: Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium, *Climate Dynamics*, 27, 165–184, 2006.
- Goosse, H., Crespin, E., de Montety, A., Mann, M., Renssen, H., and Timmermann, A.: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation, *Journal of Geophysical Research: Atmospheres*, 115, 2010.
- Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M. E., Renssen, H., Sallaz-Damaz, Y., and Shindell, D.: The role of forcing and internal dynamics in explaining the “Medieval Climate Anomaly”, *Climate dynamics*, 39, 2847–2866, 2012a.
- Goosse, H., Guiot, J., Mann, M. E., Dubinkina, S., and Sallaz-Damaz, Y.: The Medieval Climate Anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation, *Global and Planetary Change*, 84, 35–47, 2012b.
- Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., Gong, D., Kaufman, D. S., Nnamchi, H. C., Quaas, J., Rivera, J. A., Sathyendranath, S., Smith, S. L., Trewin, B., von Shuckmann, K., and Vose, R. S.: Changing State of the Climate System, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M.,

- Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., chap. 2, Cambridge University Press, 2021.
- Guttman, N. B.: A sensitivity analysis of the palmer hydrologic drought index 1, *JAWRA Journal of the American Water Resources Association*, 27, 797–807, 1991.
- Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The Last Millennium Climate Reanalysis project: Framework and first results, *Journal of Geophysical Research: Atmospheres*, 121, 6745–6764, 2016.
- Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, 129, 2776–2790, 2001.
- Hansen, J., Sato, M., Russell, G., and Kharecha, P.: Climate sensitivity, sea level and atmospheric carbon dioxide, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371, 20120294, 2013.
- Hargreaves, J. and Annan, J.: On the importance of paleoclimate modelling for improving predictions of future climate change, *Climate of the Past*, 5, 803–814, 2009.
- Hargreaves, J., Abe-Ouchi, A., and Annan, J.: Linking glacial and future climates through an ensemble of GCM simulations, *Climate of the Past*, 3, 77–87, 2007.
- Hegerl, G. C., Crowley, T. J., Hyde, W. T., and Frame, D. J.: Climate sensitivity constrained by temperature reconstructions over the past seven centuries, *Nature*, 440, 1029–1032, 2006.
- Kalman, R. E.: A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, 82, 35–45, 1960.
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S., Danabasoglu, G., Edwards, J., et al.: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change

- in the presence of internal climate variability, *Bulletin of the American Meteorological Society*, 96, 1333–1349, 2015.
- King, J., Anchukaitis, K., Allen, K., Vance, T., and Hessler, A.: Trends and variability in the Southern Annular Mode over the Common Era, in review, 2022.
- King, J. M., Anchukaitis, K. J., Tierney, J. E., Hakim, G. J., Emile-Geay, J., Zhu, F., and Wilson, R.: A data assimilation approach to last millennium temperature field reconstruction using a limited high-sensitivity proxy network, *Journal of Climate*, pp. 1–64, 2021.
- Kutzbach, J. E., He, F., Vavrus, S. J., and Ruddiman, W. F.: The dependence of equilibrium climate sensitivity on climate state: Applications to studies of climates colder than present, *Geophysical Research Letters*, 40, 3721–3726, 2013.
- Liu, H., Liu, Z., and Lu, F.: A Systematic Comparison of Particle Filter and EnKF in Assimilating Time-Averaged Observations, *Journal of Geophysical Research: Atmospheres*, 122, 13–155, 2017.
- Mairesse, A., Goosse, H., Mathiot, P., Wanner, H., and Dubinkina, S.: Investigating the consistency between proxy-based reconstructions and climate models using data assimilation: a mid-Holocene case study, *Climate of the Past*, 9, 2741–2757, 2013.
- Malevich, S. B., Vetter, L., and Tierney, J. E.: Global Core Top Calibration of  $\delta^{18}\text{O}$  in Planktic Foraminifera to Sea Surface Temperature, *Paleoceanography and Paleoclimatology*, 34, 1292–1315, 2019.
- Matsikaris, A., Widmann, M., and Jungclauss, J. H.: On-line and off-line data assimilation in palaeoclimatology: a case study, *Climate of the Past*, 11, 81–93, 2015.
- Morales, M. S., Cook, E. R., Barichivich, J., Christie, D. A., Villalba, R., LeQuesne, C., Srur, A. M., Ferrero, M. E., González-Reyes, Á., Couvreur, F., et al.: Six hundred years of South American tree rings reveal an increase in severe hydroclimatic events since mid-20th century, *Proceedings of the National Academy of Sciences*, 117, 16 816–16 823, 2020.

- Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufman, D. S., Lücke, L., Rehfeld, K., et al.: Consistent multi-decadal variability in global temperature reconstructions and simulations over the Common Era, *Nature geoscience*, 12, 643, 2019.
- Osman, M. B., Tierney, J. E., Zhu, J., Tardif, R., Hakim, G. J., King, J., and Poulsen, C. J.: Globally resolved surface temperatures since the Last Glacial Maximum, *Nature*, 599, 239–244, 2021.
- Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G.: Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model, *Bulletin of the American Meteorological Society*, 97, 735–754, 2016.
- Overpeck, J. T., Otto-Bliesner, B. L., Miller, G. H., Muhs, D. R., Alley, R. B., and Kiehl, J. T.: Paleoclimatic evidence for future ice-sheet instability and rapid sea-level rise, *Science*, 311, 1747–1750, 2006.
- Parsons, L. A., Amrhein, D. E., Sanchez, S. C., Tardif, R., Brennan, M. K., and Hakim, G. J.: Do Multi-Model Ensembles Improve Reconstruction Skill in Paleoclimate Data Assimilation?, *Earth and Space Science*, 8, e2020EA001467, 2021.
- Perkins, W. A. and Hakim, G. J.: Reconstructing paleoclimate fields using online data assimilation with a linear inverse model, *Climate of the Past*, 13, 421–436, 2017.
- Rice, J. L., Woodhouse, C. A., and Lukas, J. J.: Science and Decision Making: Water Management and Tree-Ring Data in the Western United States, *JAWRA Journal of the American Water Resources Association*, 45, 1248–1259, 2009.
- Rohling, E., Sluijs, A., Dijkstra, H., Köhler, P., van de Wal, R., von der Heydt, A., Beerling, D., Berger, A., Bijl, P., Crucifix, M., et al.: Making sense of palaeoclimate sensitivity, *Nature*, 491, 683–691, 2012.

- Rohling, E. J., Marino, G., Foster, G. L., Goodwin, P. A., Von der Heydt, A. S., and Köhler, P.: Comparing climate sensitivity, past and present, *Annual Review of Marine Science*, 10, 261–288, 2018.
- Schmidt, G., Annan, J., Bartlein, P., Cook, B., Guilyardi, E., Hargreaves, J., Harrison, S., Kageyama, M., LeGrande, A., Konecky, B., et al.: Using paleo-climate comparisons to constrain future projections in CMIP5., *Climate of the Past Discussions*, 9, 2013.
- Schmidt, G. A.: Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change, *Journal of Quaternary Science*, 25, 79–87, 2010.
- Sherwood, S., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., et al.: An assessment of Earth's climate sensitivity using multiple lines of evidence, *Reviews of Geophysics*, 58, e2019RG000678, 2020.
- Snyder, C. W.: The value of paleoclimate research in our changing climate, *Climatic Change*, 100, 407–418, 2010.
- Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of time-averaged pseudoproxies for climate reconstruction, *Journal of Climate*, 27, 426–441, 2014.
- Steiger, N. J., Steig, E. J., Dee, S. G., Roe, G. H., and Hakim, G. J.: Climate reconstruction using data assimilation of water isotope ratios from ice cores, *Journal of Geophysical Research: Atmospheres*, 122, 1545–1568, 2017.
- Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, *Scientific Data*, 5, <https://doi.org/10.1086/sdata.2018.86>, 2018.
- Stevenson, S., Otto-Bliesner, B., Brady, E., Nusbaumer, J., Tabor, C., Tomas, R., Noone, D., and Liu, Z.: Volcanic eruption signatures in the isotope-enabled last millennium ensemble, *Paleoceanography and Paleoclimatology*, 34, 1534–1552, 2019.

- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling., *Climate of the Past*, 15, 1251–1273, 2019.
- Tierney, J. E. and Tingley, M. P.: A Bayesian, spatially-varying calibration model for the TEX86 proxy, *Geochimica et Cosmochimica Acta*, 127, 83–106, 2014.
- Tierney, J. E. and Tingley, M. P.: BAYSPLINE: A new calibration for the alkenone paleothermometer, *Paleoceanography and Paleoclimatology*, 33, 281–301, 2018.
- Tierney, J. E., Malevich, S. B., Gray, W., Vetter, L., and Thirumalai, K.: Bayesian calibration of the Mg/Ca paleothermometer in planktic foraminifera, *Paleoceanography and Paleoclimatology*, 34, 2005–2030, 2019.
- Tierney, J. E., Poulsen, C. J., Montañez, I. P., Bhattacharya, T., Feng, R., Ford, H. L., Hönisch, B., Inglis, G. N., Petersen, S. V., Sagoo, N., et al.: Past climates inform our future, *Science*, 370, eaay3701, 2020a.
- Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., and Poulsen, C. J.: Glacial cooling and climate sensitivity revisited, *Nature*, 584, 569–573, 2020b.
- Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J.: An efficient forward model of the climate controls on interannual variation in tree-ring width, *Climate Dynamics*, 36, 2419–2439, 2011.
- Valler, V., Franke, J., and Brönnimann, S.: Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation, *Climate of the Past*, 15, 1427–1441, <https://doi.org/10.5194/cp-15-1427-2019>, 2019.
- Van der Schrier, G., Jones, P., and Briffa, K.: The sensitivity of the PDSI to the Thornthwaite and Penman-Monteith parameterizations for potential evapotranspiration, *Journal of Geophysical Research: Atmospheres*, 116, 2011.

- Van Leeuwen, P. J.: Particle filtering in geophysical systems, *Monthly Weather Review*, 137, 4089–4114, 2009.
- Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R<sup>2</sup>D<sup>2</sup>) bias correction, *Hydrology and Earth System Sciences*, 22, 3175–3196, 2018.
- Wang, Q. and Robertson, D.: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, *Water Resources Research*, 47, 2011.
- Wang, Q., Shrestha, D. L., Robertson, D., and Pokhrel, P.: A log-sinh transformation for data normalization and variance stabilization, *Water Resources Research*, 48, 2012.
- Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, 130, 1913–1924, 2002.
- Wikle, C. K. and Berliner, L. M.: A Bayesian tutorial for data assimilation, *Physica D: Nonlinear Phenomena*, 230, 1–16, 2007.
- Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D’arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., et al.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, *Quaternary Science Reviews*, 134, 1–18, 2016.
- Zhao, T., Bennett, J. C., Wang, Q., Schepen, A., Wood, A. W., Robertson, D. E., and Ramos, M.-H.: How suitable is quantile mapping for postprocessing GCM precipitation forecasts?, *Journal of Climate*, 30, 3185–3196, 2017.
- Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., and Anchukaitis, K. J.: Resolving the differences in the simulated and reconstructed temperature response to volcanism, *Geophysical Research Letters*, 47, e2019GL086908, 2020a.
- Zhu, F., Emile-Geay, J., Anchukaitis, K. J., Hakim, G. J., Wittenberg, A. T., Morales, M. S., Toohey, M., and King, J.: A re-appraisal of the ENSO response to volcanism with paleoclimate data assimilation, *Nature communications*, 13, 1–9, 2022.

- Zhu, J., Liu, Z., Brady, E., Otto-Bliesner, B., Zhang, J., Noone, D., Tomas, R., Nussbaumer, J., Wong, T., Jahn, A., et al.: Reduced ENSO variability at the LGM revealed by an isotope-enabled Earth system model, *Geophysical Research Letters*, 44, 6984–6992, 2017.
- Zhu, J., Poulsen, C. J., and Otto-Bliesner, B. L.: High climate sensitivity in CMIP6 model not supported by paleoclimate, *Nature Climate Change*, 10, 378–379, 2020b.
- Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Gettelman, A., Bacmeister, J. T., Neale, R. B., Poulsen, C. J., Shaw, J. K., McGraw, Z. S., and Kay, J. E.: LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2, *Journal of Advances in Modeling Earth Systems*, p. e2021MS002776, 2021a.
- Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Poulsen, C. J., Tierney, J. E., Lofverstrom, M., and DiNezio, P.: Assessment of equilibrium climate sensitivity of the Community Earth System Model version 2 through simulation of the Last Glacial Maximum, *Geophysical Research Letters*, 48, e2020GL091 220, 2021b.

## C.10 Supplementary Information

### C.10.1 Example 1: Northern Hemisphere Summer Temperatures over the Last Millennium

```
1 %% Example 1: NTREND Assimilation
2
3 % Reset the random number generator so all examples are reproducible
4 rng('default');
5
6 %% gridfile: Organize climate model output
7
8 % List of climate model output files
9 outputFile1 = 'b.e11.BLMTRC5CN.f19_g16.002.cam.h0.TREFHT.085001-184912.
   nc';
10 outputFile2 = 'b.e11.BLMTRC5CN.f19_g16.002.cam.h0.TREFHT.185001-200512.
   nc';
11
12 % Define metadata that spans the climate model output dataset
13 lat = ncread(outputFile1, 'lat');
14 lon = ncread(outputFile1, 'lon');
15 time = datetime(850,1,15):calmonths(1):datetime(2005,12,15);
16
17 metadata = gridMetadata('lat', lat, 'lon', lon, 'time', time);
18 metadata = metadata.addAttributes('Units', 'Kelvin', 'Model', 'CESM 1.0
   ');
19
20 % Initialize a new, empty gridfile to catalogue the dataset
21 modelOutput = gridfile.new('Temperature-CESM', metadata);
22
23 % Catalogue the source files for the dataset in the gridfile
24 dimensionOrder = ["lon","lat","time"];
25 output1Metadata = metadata.index('time', 1:12000);
26 output2Metadata = metadata.index('time', 12001:13872);
27
```

```
28 modelOutput.add('netcdf', outputFile1, "TREFHT", dimensionOrder,
    output1Metadata);
29 modelOutput.add('netcdf', outputFile2, "TREFHT", dimensionOrder,
    output2Metadata);
30
31 % Convert loaded data from Kelvin to Celsius
32 modelOutput.transform('plus', -273.15);
33
34
35 %% gridfile: Organize climate proxy records
36
37 % File holding the proxy record dataset
38 proxyFile = 'ntrend.mat';
39
40 % Define metadata for the proxy record dataset
41 info = load(proxyFile, 'years', 'site_IDs', 'lons', 'lats', 'seasons');
42 site = [info.site_IDs, info.lats, info.lons, info.seasons];
43 proxyMetadata = gridMetadata('site', site, 'time', info.years);
44
45 % Catalogue the proxy record dataset in a gridfile
46 proxies = gridfile.new('ntrend', proxyMetadata);
47 proxies.add('mat', proxyFile, 'crn', ["time" "site"], proxyMetadata);
48
49 % Indicate that -999 is a fill value and should be converted to NaN
50 proxies.fillValue(-999);
51
52
53 %% State vector: Design and build a state vector ensemble
54
55 % Initialize a state vector and label the object
56 sv = stateVector('NTREND Assimilation');
57
58 % Initialize variables that are:
59 % 1. Used to run the proxy forward models, or
60 % 2. Reconstruction targets.
61 proxyNames = proxyMetadata.site(:,1);
```

```

62 sv = sv.add(proxyNames, modelOutput);
63 sv = sv.add(["T", "T_index"], modelOutput);
64
65 % Specify that ensemble members will be selected from each year of
    model output
66 % (Note that we are only using January as a reference month for each
    year.
67 % Later steps will specify the monthly means used in the assimilation).
68 january = month(metadata.time) == 1;
69 sv = sv.design(0, 'time', 'ensemble', january);
70
71 % Design the reconstruction targets to use data north of 35N and a
    seasonal
72 % mean over June, July, and August. Also use a latitude-weighted
    spatial
73 % mean for the reconstructed extratropical temperature index.
74 extratropical = metadata.lat > 35;
75 JJA = [5 6 7];
76 latWeights = cosd(metadata.lat(extratropical));
77
78 sv = sv.design(["T","T_index"], 'lat', [], extratropical);
79 sv = sv.mean( ["T","T_index"], 'time', JJA);
80 sv = sv.weightedMean("T_index", ["lat" "lon"], {latWeights, []});
81
82 % Design the proxy variables to use the site-specific seasonal
    temperature
83 % mean from the model grid closest to the proxy site.
84 nProxies = numel(proxyNames);
85 proxyCoordinates = str2double( proxyMetadata.site(:,2:3) );
86 nearestGrids = dash.closestLatLon(proxyCoordinates, metadata.lat,
    metadata.lon);
87
88 for p = 1:nProxies
89     latIndex = metadata.lat == nearestGrids(p,1);
90     lonIndex = metadata.lon == nearestGrids(p,2);
91     sv = sv.design(p, ["lat","lon"], [], {latIndex, lonIndex});

```

```

92
93     season = strsplit(proxyMetadata.site(p,4), ',');
94     season = str2double(season) - 1;
95     sv = sv.mean(p, 'time', season);
96 end
97
98 % Build a state vector ensemble with 1000 members and save it to file
99 [ens, ensMeta] = sv.build(1000, 'file', 'ntrend-ensemble');
100
101
102 %% PSM: Implement forward models for the NTREND sites and estimate
    proxy values
103
104 % Load coefficients of linear forward models
105 coeffs = load('ntrend-forward-model-coefficients');
106 slopes = coeffs.slopes;
107 intercepts = coeffs.intercepts;
108
109 % Design a simple linear forward models for each proxy record
110 forwardModels = cell(nProxies, 1);
111 for p = 1:nProxies
112     forwardModels{p} = PSM.linear(slopes(p), intercepts(p));
113
114     % Indicate the row in the state vector ensemble that holds the
        seasonal
115     % temperature means needed to run the forward model.
116     row = ensMeta.findRows( proxyNames(p) );
117     forwardModels{p} = forwardModels{p}.useRows(row);
118 end
119
120 % Compute proxy estimates by running the forward models on the ensemble
121 Ye = PSM.estimate(forwardModels, ens);
122
123
124 %% Kalman Filter: Implement a Kalman Filter and generate a
    reconstruction

```

```

125
126 % Initialize a Kalman Filter
127 kf = kalmanFilter('NTREND Assimilation');
128
129 % Indicate the prior. This will be the variables in the ensemble that
130 % correspond to reconstruction targets.
131 [ens, ensMeta] = ens.useVariables(["T","T_index"]);
132 kf = kf.prior(ens);
133
134 % Specify the proxy records and estimates used in assimilation
135 Y = proxies.load;
136 kf = kf.observations(Y);
137 kf = kf.estimates(Ye);
138
139 % Also specify proxy error variances
140 R = load('ntrend-error-variances').R;
141 kf = kf.uncertainties(R);
142
143 % Implement covariance localization with a radius of 20000 km
144 coordinates = ensMeta.latlon;
145 radius = 20000; % km
146 [wloc, yloc] = dash.localize.gaspariCohn2D(radius, coordinates,
      proxyCoordinates);
147 kf = kf.localize(wloc, yloc);
148
149 % Also calculate a latitude-weighted, spatial mean temperature index
      from
150 % the updated spatial field. This will return the full posterior
      ensemble
151 % for the index, without needing to save the (very large) posterior
152 % ensemble of the full spatial temperature field.
153 latWeights = cosd(coordinates(:,1));
154 kf = kf.index('T_index2', 'mean', latWeights);
155
156 % Save posterior percentiles and variance, rather than the full
      posterior.

```

```
157 kf = kf.percentiles(0:5:100);
158 kf = kf.variance(true);
159
160 % Run the Kalman Filter.
161 output = kf.run;
162
163 % Extract reconstruction outputs, and regrid the spatial field to match
    the
164 % dimensions of the initial climate model output
165 [T, Tmeta] = ensMeta.regrid("T", output.Amean, ["lat","lon"]);
166 Tvar = ensMeta.regrid("T", output.Avar, ["lat","lon"]);
167 index1 = output.Amean(end,:);
168 index2 = output.index.T_index2;
169
170 % Save for post-processing and visualization
171 save('ntrend-reconstruction.mat', 'T', 'Tvar', 'Tmeta', 'index1', 'index2');
172
173
174 %% Optimal sensor
175
176 % Initialize an optimal sensor for the NTREND assimilation
177 os = optimalSensor('NTREND sensors');
178
179 % Use the summer-temperature spatial field as the prior. Also specify
    the
180 % proxy estimates and error variances used in the analysis.
181 ens = ens.useVariables('T');
182 os = os.prior(ens);
183 os = os.estimate(Ye, R);
184
185 % Use the spatial mean of the extratropical, summer-temperature spatial
186 % field as the sensor metric
187 latWeights = latWeights(1:end);
188 os = os.metric('mean', latWeights);
189
```

```
190 % Run the sensor to evaluate the ability of each record to reduce
      uncertainty
191 proxyPower = os.evaluate;
192
193 % Save for analysis and visualization
194 save('ntrend-optimal-sensor', 'proxyPower', 'proxyMetadata');
```

### C.10.2 Example 2: Global Sea Level Pressures from the Last Glacial Maximum to Present

```

1 %% Example 2: LGM Assimilation
2
3 %% gridfile: Organize climate model output
4
5 % Climate model output
6 slpFile = 'PSL_iCESM_cam_50yrMonthlyAverages_LGMtoPresent.nc';
7 sstFile = 'TEMP_iCESM_pop_50yrMonthlyAverages_LGMtoPresent.nc';
8 d18OFile = 'R180_iCESM_pop_50yrMonthlyAverages_LGMtoPresent.nc';
9
10 % Define metadata for the rectilinear, atmospheric model output
11 lat = ncread(slpFile, 'lat');
12 lon = ncread(slpFile, 'lon');
13 time = ncread(slpFile, 'time');
14 camMetadata = gridMetadata('lat', lat, 'lon', lon, 'time', time);
15
16 % Define metadata for the tripolar, ocean model output
17 lat = ncread(outputFile, 'lat');
18 lon = ncread(outputFile, 'lon');
19 site = [lat(:), lon(:)];
20 popMetadata = gridMetadata('site', site, 'time', time);
21
22 % Initialize a gridfile object for each variable
23 slp = gridfile.new('SLP', camMetadata);
24 sst = gridfile.new('SST', popMetadata);
25 d18O = gridfile.new('d18O', popMetadata);
26
27 % Catalogue the output files
28 slp.add('netcdf', slpFile, "PSL", ["lon", "lat", "time"],
        camMetadata);
29 sst.add('netcdf', sstFile, "TEMP", ["site", "site", "time"],
        popMetadata);
30 d18O.add('netcdf', d18OFile, "R180", ["site", "site", "time"],
        popMetadata);

```

```

31
32
33 %% gridfile: Organize climate proxy records
34
35 % Files holding the proxy record dataset and metadata
36 proxyDataFile = 'proxies.txt';
37 proxyMetaFile = 'proxy-metadata.txt';
38
39 % Define metadata for the proxy record dataset
40 info = readtable(proxyMetaFile);
41 site = [info.ID, info.lat, info.lon, info.proxyType, info.species];
42 proxyMetadata = gridMetadata('site', site, 'time', info.time);
43
44 % Catalogue the proxy record dataset in a gridfile
45 proxies = gridfile.new('proxies', metadata);
46 proxies.add('text', proxyDataFile, ["time","site"], 'NumHeaderLines',
47     1);
48
49 %% Design and build state vector ensemble
50
51 % Initialize a state vector and label the object
52 sv = stateVector('LGM Assimilation');
53
54 % Initialize variables that are:
55 % 1. Reconstruction targets
56 %     - Sea level pressure (SLP)
57 % 2. Used to run the proxy forward models
58 %     - Monthly sea surface temperature (SST)
59 %     - Mean-annual d18O_sw
60 sv = sv.add(["SLP","SST","d18O"], [slp;sstFile;d18O]);
61
62 % Specify that ensemble members will be selected from each 50 year bin
63 % of
64 % monthly values
65 january = camMetadata.time(:,2)==1;

```

```

65 sv = sv.design(0, 'time', 'ensemble', january);
66
67 % Use 50-yr annual means for SLP and d18Osw. Include all monthly 50-
    year
68 % bins for SST
69 months = (0:11)';
70 sv = sv.mean(["SLP","d18Oc"], 'time', months);
71 sv = sv.sequence("SST", 'time', months, months);
72
73 % Build the state vector sequentially using all available 50-year bins
74 [ens, ensMeta] = sv.build('all', 'sequential', true, 'file', 'lgm-
    ensemble');
75
76 % There are 16 50-year bins for each of the nine 3,000 year intervals.
77 % Use the sets of 16 ensemble members to build an evolving ensemble for
    the
78 % nine intervals.
79 members = 1:ens.nMembers;
80 members = reshape(members, 16, 9);
81 ens = ens.evolver(members);
82
83
84 %% PSM: Implement forward models for the proxy sites and estimate proxy
    values
85
86 % Download proxy forward models from Github
87 PSM.download('bayfox');
88 PSM.download('bayspline');
89
90 % Design a forward model for each proxy record
91 nProxies = numel(proxyInfo.ID);
92 forwardModels = cell(nProxies, 1);
93 for p = 1:nProxies
94
95     % Get the proxy ID, location, and type
96     ID             = proxyMetadata.site(p,1);

```

```

97     coordinates = proxyMetadata.site(p,2:3);
98     type         = proxyMetadata.site(p,4);
99
100    % Use either a UK'37 forward model, which requires monthly SSTs
101    if type == "uk37"
102        SST = ensMeta.closestLatLon("SST", coordinates);
103        model = PSM.bayspline;
104        model = model.useRows(SST);
105
106    % Or a d180_c model, which is calibrated to different foraminiferal
107        species...
108    elseif type == "d180c"
109        species = proxyMetadata.site(p,5);
110        model = PSM.bayfox(species);
111
112    % ...and which requires monthly SSTs and annual d180_sw
113    SST = ensMeta.closestLatLon("SST", coordinates);
114    d180 = ensMeta.closestLatLon("d180", coordinates);
115    mod = model.useRows([SST;d180]);
116
117    end
118
119    % Label and save the model for each proxy
120    model = model.label(ID);
121    forwardModels{p} = model;
122 end
123
124 % Run the forward models on the ensemble to compute proxy estimates and
125 % proxy error variances
126 [Ye, R] = PSM.estimate(forwardModels, ens);
127
128 %% Kalman Filter
129
130 % Initialize and label a kalman filter
131 kf = kalmanFilter('LGM Assimilation');

```

```
132 % Provide, the prior, proxy records, estimates, and error variances
133 Y = proxies.load;
134
135 kf = kf.prior(ens);
136 kf = kf.observations(Y);
137 kf = kf.estimated(Ye);
138 kf = kf.uncertainties(R);
139
140 % Run the Kalman filter. To conserve memory, return the mean and
      variance
141 % of the posterior ensemble, rather than the complete ensemble
142 kf = kf.variance(true);
143 output = kf.run;
144
145 % Regrid the SLP reconstruction target
146 [SLP, SLPmeta] = ensMeta.regrid("SLP", output.Amean);
147 SLPvar = ensMeta.regrid("SLP", output.Avar);
148
149 % Save for visualization
150 save('lgm-reconstruction', 'SLP', 'SLPmeta', 'SLPvar');
151
152
153 %% Proxy validation: Run the proxy system models on the posterior
      ensemble
154
155 % Run the proxy forward models on the posterior ensemble
156 Ypost = PSM.estimate(forwardModels, output.Amean);
157
158 % Save for post-processing and visualization
159 save('lgm-proxy-validation', 'Ypost', 'proxyMetadata');
```

Table C.1: Proxy forward-models currently supported by DASH.

<b>Model</b>	<b>Description</b>	<b>Citation or Authors</b>	<b>Github Repository</b>
BayFOX	Bayesian model of planktic foraminiferal $\delta^{18}O_c$	Malevich et al. (2019)	jesstierney/bayfoxm
BayMAG	Bayesian model of planktic foraminiferal Mg/Ca	Tierney et al. (2019)	jesstierney/BAYMAG
BaySPAR	Bayesian model for TEX'86	Tierney and Tingley (2014)	jesstierney/BAYSPAR
BaySPLINE	Bayesian model for UK'37	Tierney and Tingley (2018)	jesstierney/BAYSPLINE
Multi-variate Linear	General multi-variate linear forward models		DASH built-in
PDSI	Palmer Drought-Severity Index estimator	Dave Meko, Jonathan King	JonKing93/pdsi
PRYSM Cellulose	Cellulose $\delta^{18}O$	Dee et al. (2015)	sylvia-dee/PRYSM
PRYSM Coral	Coral $\delta^{18}O$	Dee et al. (2015)	sylvia-dee/PRYSM
PRYSM Ice-Core	Ice-core $\delta^{18}O$	Dee et al. (2015)	sylvia-dee/PRYSM
PRYSM Speleothem	Speleothem $\delta^{18}O$	Dee et al. (2015)	sylvia-dee/PRYSM
VS-Lite	Vaganov-Shashkin Lite model of tree-ring width	Tolwinski-Ward et al. (2011)	suztolwinskiward/vslite

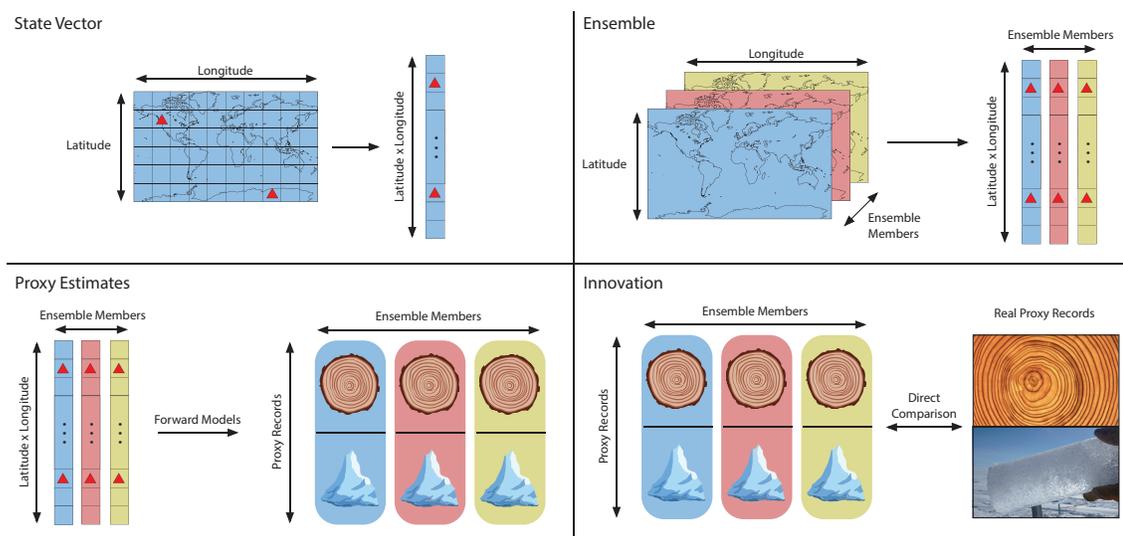


Figure C.1: Illustration of common tasks and vocabulary for paleoclimate data assimilation. Top left: Gridded climate model output is reshaped into a *state vector*. Red triangles indicate the locations of proxy records. Top right: Multiple climate model outputs are reshaped into state vectors and concatenated into an *ensemble*. Bottom left: Forward models are applied to each state vector and used to generate *proxy estimates* for each proxy record. Bottom right: Proxy estimates are compared directly to the real proxy records. The difference between the estimates and the real records is the *innovation*.



Figure C.2: Flowchart illustrating DASH components and their uses within the context of paleoclimate data assimilation workflows.

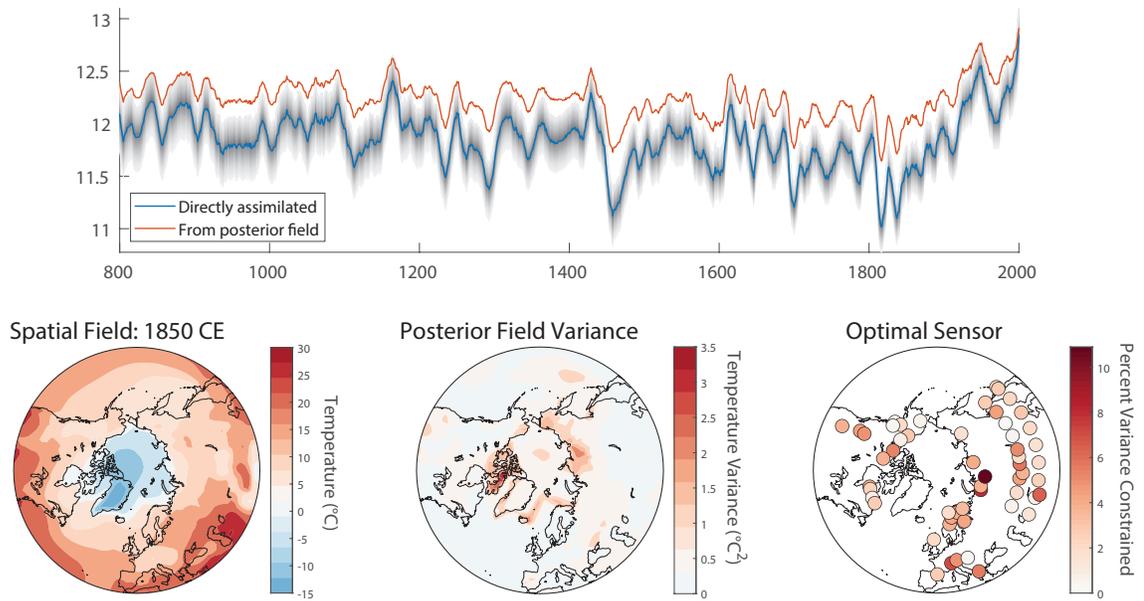


Figure C.3: Results from Example 1, the NTREND assimilation. Top: Reconstructed mean extratropical summer (June-August) temperatures. The blue line shows the reconstructed index when the index is assimilated directly in the state vector. The red line shows the index calculated from the posterior spatial field. Grey shading indicates the 5-95% confidence level for Index 1. Lower left: The reconstructed summer-temperature spatial field in the year 1850 CE. Lower center: The variance of the posterior spatial field in the year 1850 CE. High variance indicates greater uncertainty in the reconstructed spatial field. Lower right: Results of the optimal sensor analysis. Circles indicate the locations of the NTREND tree-ring records. The color of each circle indicates the percent variance of the reconstructed index that is constrained by assimilating each NTREND site individually.

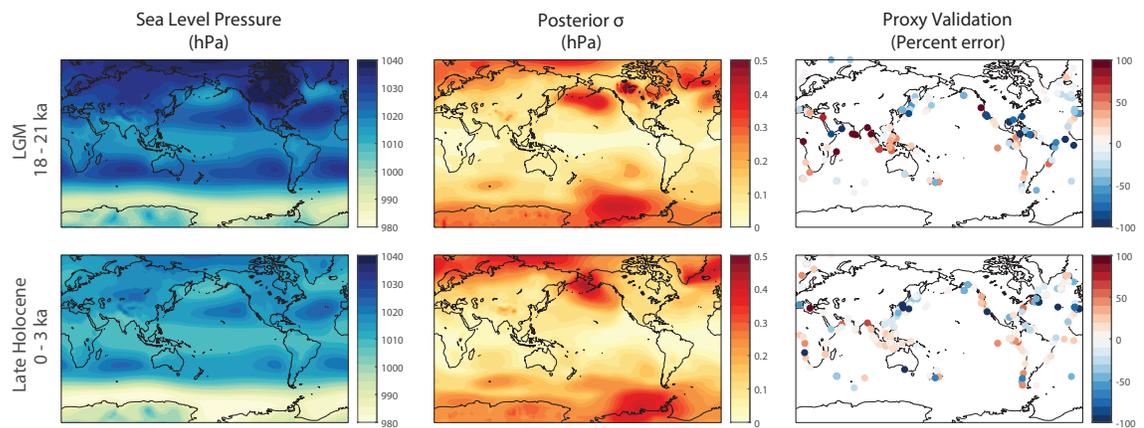


Figure C.4: Results from Example 2, the LGM assimilation. Upper row shows results for the Last Glacial Maximum (18-21 ka); lower row shows results for the most recent 3,000-year interval (0-3 ka). From left to right, columns display reconstructed sea level pressure fields (hPa), the standard deviation across the posterior ensembles for each reconstructed field (hPa), and the percent errors from the proxy validations.